THE ANNALS *of* MATHEMATICAL STATISTICS

(FOUNDED BY H. C. CARVER)

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. X, No. 2 — June, 1939

THE ANNALS OF MATHEMATICAL STATISTICS

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I–IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

Subscriptions, renewals, orders for back numbers and other business communications should be sent to A. T. Craig, University of Iowa, Iowa City, Iowa.

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

THE LIMITS OF A DISTRIBUTION FUNCTION IF TWO EXPECTED VALUES ARE GIVEN

By R. v. Mises

In a very interesting paper[1] A. Wald dealt with the following generalization of a problem started by Markoff and Tchebycheff: Denote by X a random variable, by $P(t)$ the probability of $|X| < t$ and by M_r the absolute moment of order r or the expected value of $|X|^r$; what is the sharp lower limit (limes inferior) of $P(t)$ for any point t, if $M_\alpha, M_\beta, M_\gamma, \cdots$ are given? Wald outlines an ingenious method for the general case of n given moments and adds the complete solution for the case $n = 2$. I wish to show in the following lines that the results for $n = 2$ can be deduced both in a more general and less complicated manner. Instead of two different powers of $|X|$, I shall admit largely arbitrary functions of X and I shall get the solution by a more intuitive way. Moreover the upper limit of $P(t)$ will be found too. It seems to me that my method will be applicable also to certain cases with $n > 2$.

1. **The Problem.** Without loss of generality we can restrict ourselves to a non-negative random variable X. Let $x(X)$ and $y(X)$ be two increasing functions of X with $x(0) = y(0) = 0$. We suppose that the curve defined in a Cartesian co-ordinate system by

$$x = x(t), \qquad y = y(t) \tag{1}$$

is one which is convex downwards, i.e. the slope of its chords is increasing if the co-ordinates of one or both extreme points of the chord increase (see Fig. 1). This condition is fulfilled, for instance, if $x = t^r$, $y = t^s$ and $s > r > 0$ where the indexes r, s are not necessarily integers. Another example is $x = t, y = t^2/1 + t$; here, however, the ratio y/x is restricted to values between 0 and 1. In a third class of examples as $x = t/1 + t$, $y = t^3/(1 + t)^3$ the curve corresponding to (1) ends at a finite point.

The probability of the inequality $X < t$ will be designated by $P(t)$, the probability of $X > t$ by $\overline{P}(t)$. The sum of $P(t)$ and $\overline{P}(t)$ is equal to 1 excepting the points t associated with a finite probability. But in any case the upper limit of $P(t)$ and the lower limit of $\overline{P}(t)$ give the sum 1.

The expected values of $x(X)$ and $y(X)$ can be defined by means of $P(t)$ or $\overline{P}(t)$

$$\begin{aligned} a &= \int_0^\infty x(t)\, dP(t) = -\int_0^\infty x(t)\, d\overline{P}(t); \\ b &= \int_0^\infty y(t)\, dP(t) = -\int_0^\infty y(t)\, d\overline{P}(t). \end{aligned} \tag{2}$$

[1] *Ann. Math. Statist.* Vol. 9 (1938) pp. 244–255.

We suppose that the values of a and b are given in a suitable manner and we ask for the lower limit of $P(t)$ and $\overline{P}(t)$ at any point t. In other words, we try to find two functions $l(t)$ and $\bar{l}(t)$ so that for all distributions associated with the given values of a and b we have

$$P(t) \geqq l(t), \qquad \overline{P}(t) \geqq \bar{l}(t), \tag{3}$$

but that these inequalities are not valid, if $l(t)$ and $\bar{l}(t)$ are replaced by higher values. In Fig. 1 K is the curve defined by (1) and C the point with co-ordinates a, b.

We can give a more intuitive interpretation to our problem by imagining a mass distribution instead of a probability or frequency distribution. In fact, if the mass of magnitude 1 is spread along the curve K in such a way that $P(t)$ designates the sum (or integral) of masses lying to the left of the point $x(t)$, $y(t)$, then the point C will be the centre of gravity (centre of mass) of the whole mass system. By the way, it follows that C must be situated on the inner side of the convex curve K. Our question can now be stated as follows:

A mass of size 1 is distributed along a given convex curve and has its centre of gravity in a given point C. What is the least possible value of mass lying to the left or to the right of any point $x(t)$, $y(t)$ of the curve?

2. **Restriction of Distributions to be Considered.** The essential difficulty of our problem lies in the fact that in order to find the limits $l(t)$ and $\bar{l}(t)$ all conceivable forms of distribution functions $P(t)$ and $\overline{P}(t)$ must be taken into account. Let us now see how the field of distributions can be restricted in a decisive manner.

Two mass systems with the same total mass and the same centre of gravity will be called "equivalent systems". Then the following corollary can be stated:

If a mass system with the distribution functions P, $\overline{P}$ and a point M with co-ordinates $x(t)$, $y(t)$ are given, we can always find an equivalent system consisting of three particles or masspoints: a mass m_1 at M_1 to the left of M, a mass m at M itself and a mass m_2 at M_2 to the right of M, so that

$$m_1 \leqq P(t), \qquad m_2 \leqq \overline{P}(t). \tag{4}$$

This proposition enables us, in asking for the lower limit values $l(t)$, $\bar{l}(t)$ at M to confine ourselves to the consideration of a special class of three-point systems and to disregard all other kinds of distributions.

In order to prove the corollary we make use of the well known laws of elementary statics. According to these laws all masses lying to the left of M (in the given system) can be replaced by a single mass of same size fixed in their centre of gravity C_2 (Fig. 1). This centre is situated in the domain between the curve K and the chord OM. The straight line MC_1 has one and only one second point of intersection M_1 with K. Any mass at C_1 can be decomposed into two masses, one of them of magnitude m_1 lying at M_1, the other of size m' at M.

In an analogous manner starting with the masses lying to the right of M in the given system, m_2, m'' and M_2 can be found. It is evident that m_1 can not exceed the sum of masses which were attached to points to left of M in the original mass system. It is the same with m_2 and the masses to the right of M. If in the original system a finite mass m_0 had been attached to the point M, the value of m in the new mass system will be defined as $m = m' + m'' + m_0$.

3. **The Extreme Distributions.** Now, in order to find the limits $l(t)$ and $\bar{l}(t)$ for a certain point M, we are concerned exclusively with a two-parameter family of mass systems, each of them consisting of three masses m_1, m, m_2 at three points M_1, M, M_2. We choose as parameters the magnitude m of the mass attached to the point M and the slope of the chord joining M_1 and M_2. If m remains constant, the chord M_1M_2 (Fig. 2) passes through a fixed point C_0 on

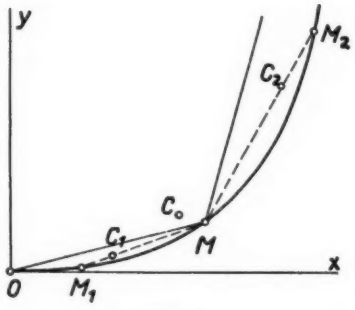

FIG. 1

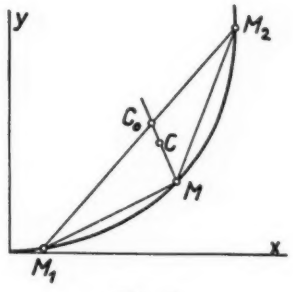

FIG. 2

the prolongation of MC where $\overline{CC_0} = \overline{MC}.m/1 - m$. The masses m_1 and m_2 vary with the direction of M_1M_2 and are determined by

$$(5) \qquad m_1 = (1 - m)\frac{\overline{C_0M_2}}{\overline{M_1M_2}}, \qquad m_2 = (1 - m)\frac{\overline{M_1C_0}}{\overline{M_1M_2}}$$

We are only interested in the least possible values of m_1 and m_2. But a convex curve for which the angle formed by its extreme tangents is not greater than 90°, has the characteristic property that the ratio of chord segments $\overline{M_1C_0} : \overline{C_0M_2}$, for an inner point C_0, is permanently increasing or decreasing when the chord turns about C_0; there is no analytical maximum or minimum. It follows that the lowest values of m_1 and m_2 can only be found in an extreme position of the chord, i.e. when M_1 coincides with O, or M_2 with the other (eventually infinite) end Q of K, or finally when one of the points M_1, M_2 coincides with M. The latter cases must be mentioned since it was one of the conditions for our three-point systems that M lies *between* M_1 and M_2. The result we have obtained until now is, that the lowest values of masses lying on one or the other side of M are to be sought in a distribution of one of the following classes: (1) The three-mass systems with one mass at M and one mass at O; (2) The three-

mass systems with one mass at M and one mass at the end Q of K; (3) The two-mass system with one mass at M.

Now we must distinguish three sorts of points M or three sections of the curve K. If we trace the chord (see Fig. 3) beginning at O and passing through C we obtain the point of intersection O' and by means of the chord QC we arrive at the point Q'. The three sections of K we have to deal with are OQ', $Q'O'$ and $O'Q$.

If M is a point of OQ' there exists a chord MM' passing through C and therefore a two-mass system with masses m, m' at M and M'. In this system the mass to the left of M is zero, thus we have $l(t) = 0$ for all these points. If we consider a three-mass system with one mass at M, one mass at O and one mass at any point M_2, the value of m_2 is equal to the ratio $\overline{CC_1}/\overline{M_2C_1}$, where C_1 is the intersection of M_2C with OM. The least value of this ratio will be reached when C_1 coincides with M. Therefore $l(t)$ is equal to the ratio $\overline{CM}/\overline{M'M}$ or equal to the mass m' of the two-mass system mentioned before.

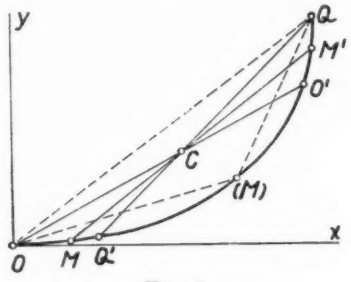

FIG. 3

Now let M be a point of the arc $Q'O'$. For such a point a two-mass system does not exist, since the straight line $M\,C$ does not meet the curve a second time. In a three-mass system $O\,M\,M_2$ the value of m is equal to $\overline{CC_1}/\overline{M_2C_1}$ as before, and the least value of this ratio is attained, if M_2 coincides with Q. It follows that $l(t)$ is equal to the ratio $\overline{CC_Q}/\overline{QC_Q}$ where C_Q denotes the point of intersection of QC with OM. In the same way we find $l(t)$ equal to $\overline{CC_O}/OC_O$, the point of intersection of OC with MQ being designated by C_O.

For a point M of the arc $O'Q$ the circumstances are the same as for the points of OQ'.

In other words the extreme distributions which furnish immediately the values of $l(t)$ and $\bar{l}(t)$ are 1) the two-mass systems MM' for all points of the arcs OQ' and $O'Q$ and 2) the three-mass system OMQ for a point of the middle section $Q'O'$. The corresponding values of l and $\bar{l}$ are to be found by the elementary laws of statics in the simplest way.

4. Results. The definite results can now be stated as follows. Our data are the functions $x(t)$, $y(t)$ and the expected values a, b.

First we compute the co-ordinates p, q of the endpoint Q, i.e. $p = x(\infty)$,

$q = y(\infty)$. If q or p and q are infinite, we only need the limit value of y/x. Then the two values t_0 and t^0 corresponding to the points O' and Q' have to be found. They are determined by the equations

$$\frac{y(t_0)}{x(t_0)} = \frac{b}{a}; \qquad \frac{y(t^0) - b}{x(t^0) - a} = \frac{b - q}{a - p}. \tag{6}$$

If t belongs to one of the intervals $t \leqq t^0$ or $t \geqq t_0$, there exists one and only one value of t' different from t and satisfying the equation

$$\frac{a - x(t)}{x(t') - x(t)} = \frac{b - y(t)}{y(t') - y(t)}. \tag{7}$$

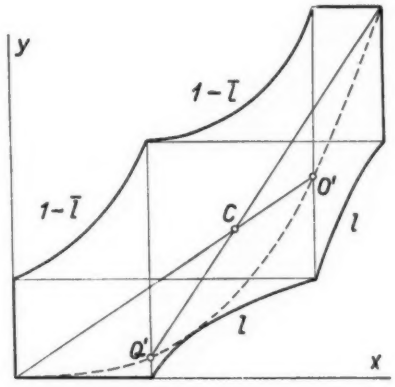

FIG. 4

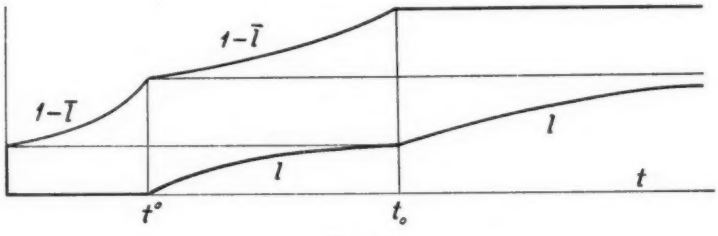

FIG. 5

The point M' with co-ordinates $x' = x(t')$, $y' = y(t')$ is the second endpoint of the chord passing through M and C. Now we have, according to the preceding considerations:

$$\text{For } t \leqq t^0: \qquad l(t) = 0, \qquad l(t) = \frac{a - x}{x' - x}$$

$$\text{(8)} \quad \text{“} \quad t^0 \leqq t \leqq t_0: \; l(t) = \frac{(a - p)(b - y) - (a - x)(b - q)}{py - qx} \qquad l(t) = \frac{ay - bx}{py - qx}$$

$$\text{“} \quad t \geqq t_0: \qquad l(t) = \frac{a - x}{x' - x}, \qquad l(t) = 0$$

The formulae are considerably simplified, if p, q are infinite. In the case of two moments given, $x = t^r$, $y = t^s$, $s > r > 0$, we have $p = \infty$, $q = \infty$, $\lim y/x = \infty$. The second equation (6) gives $x(t^0) = a$ and (8) becomes:

$$
\begin{array}{lllll}
 & \text{For } t \leqq t^0 & l(t) = 0, & & l(t) = \dfrac{a - x}{x' - x} \\
(9) & \text{“ } t^0 \leqq t \leqq t_0 & l(t) = \dfrac{x - a}{x}, & & l(t) = 0 \\
 & \text{“ } t \geqq t_0, & l(t) = \dfrac{a - x}{x' - x} & & l(t) = 0
\end{array}
$$

The values of $l(t)$ given here are in full accordance with the results published by Wald in his paper quoted above.

A great part of the numerical investigation is independent from the relation which joins x (or y) to t and is determined only by the values a, b, the curve K, i.e. the relation between x and y, and its endpoint Q. In the following example we have assumed $y = x^3$ and as endpoint $p = q = 1$. Fig. 4 shows for $a = 0, 6$, $b = 0, 4$ the three sections of the lines l and $1 - l$ according to the equations (8), but with the abscissae x. The graph of any distribution function in the interval $0 \leqq x \leqq 1$ with given first moment 0, 6 and third moment 0, 4 keeps within the space between the lines l and $1 - l$. If we now assume, e.g. $x = t^2/1 + t^2$ the abscissae x are to be transformed according to this equation and the graphs of definitive $l(t)$ and $1 - l(t)$ functions are those given in fig. 5. Any distribution functions $P(t)$ with the expected values

$$\int \frac{t^2}{1 + t^2}\, dP(t) = 0, 6 \qquad \int \left(\frac{t^2}{1 + t^2}\right)^3 dP(t) = 0, 4$$

must keep between the two limits indicated in Fig. 5. If such a function touches the upper limit in any point, it will also attain the lower limit in another point and will correspond to a two- or three-mass system.

UNIVERSITY OF ISTANBUL,
ISTANBUL, TURKEY.

CONFIDENCE LIMITS FOR CONTINUOUS DISTRIBUTION FUNCTIONS[1]

By A. Wald[2] and J. Wolfowitz

1. **Introduction.** The theory of confidence limits for unknown parameters of distribution functions has been considerably developed in recent years. This theory assumes that there is given a family F of systems of n stochastic variables $X_1(\theta_1, \cdots, \theta_k), \cdots, X_n(\theta_1, \cdots, \theta_k)$ depending upon k parameters $\theta_1, \cdots, \theta_k$ and such that the distribution function of every element of F is known.

For the case $k = 1$, for example, this theory proceeds as follows:

Denote by E an n-tuple $x_1, \cdots, x_n$ of observed values of the stochastic variables $X_1(\theta), \cdots, X_n(\theta)$ of which we know only that they constitute a system which is an element of F. E can be represented as the point $x_1, \cdots, x_n$ in an n-dimensional Euclidean space. Let there be given a positive number α, $0 < \alpha < 1$. Then to each pair E, α there is constructed a $\underline{\theta}$-interval, $[\underline{\theta}(E, \alpha), \bar{\theta}(E, \alpha)]$ with the following property: If we were to draw a sample from the system $X_1(\theta), \cdots, X_n(\theta)$, the probability is exactly α that we shall get a system of observations $E = x_1, \cdots, x_n$ such that the interval corresponding to E, α will include θ (i.e., that $\underline{\theta}(E, \alpha) \leq \theta \leq \bar{\theta}(E, \alpha)$).

In this paper we do not limit ourselves to a family of systems of n stochastic variables depending upon a finite number of parameters, but consider the family G of all systems of n stochastic variables $X_1, \cdots, X_n$ subject only to the condition that $X_1, \cdots, X_n$ are independently distributed with the same continuous distribution function.

Let E be the point in an n-dimensional Euclidean space which corresponds to the observed values $x_1, \cdots, x_n$ of the n stochastic variables $X_1, \cdots, X_n$ of which we know only that they constitute an element of the family G, i.e., that they are independently distributed with the same continuous distribution function. Let us denote their distribution function by $f(x)$; the probability that $X_i < x$ is $f(x)$, $i = 1, \cdots, n$. Let α be a number such that $0 < \alpha < 1$. To each pair E, α we shall construct two functions, $l_{E,\alpha}(x)$ and $\underline{l}_{E,\alpha}(x)$, with the following property: The probability is α that, if we were to draw a sample from the system $X_1, \cdots, X_n$, we would get a system of observations $E = x_1, \cdots, x_n$ such that $f(x)$ lies entirely between $l_{E,\alpha}(x)$ and $\underline{l}_{E,\alpha}(x)$ (i.e., that $\underline{l}_{E,\alpha}(x) \leq f(x) \leq l_{E,\alpha}(x)$ for all x). We shall call $l_{E,\alpha}(x)$ and $\underline{l}_{E,\alpha}(x)$ the upper and lower confidence limits, respectively, corresponding to the confidence coefficient α.

---

[1] Presented to the American Mathematical Society at New York, February 25, 1939.

[2] Research under a grant-in-aid from the Carnegie Corporation of New York.

All the stochastic variables considered hereafter in this paper are to have continuous distribution functions.

2. **A theorem on continuous distribution functions.** Let $f(x)$ be the continuous distribution function of a stochastic variable X whose range is from $-\infty$ to $+\infty$. Let $\delta_1(x)$ and $\delta_2(x)$ be two functions defined for $0 \leq x \leq 1$ and satisfying the following requirements:

(a) $\delta_1(x)$ and $\delta_2(x)$ are non-negative and continuous for $0 \leq x \leq 1$.

(b) $l_1(x)$ and $l_2(x)$ are monotonically non-decreasing for all x, where

$$l_1(x) \equiv f(x) + \delta_1(f(x))$$

$$l_2(x) \equiv f(x) - \delta_2(f(x)).$$

(c) There exists a number h, such that $f(h) < 1$ and $l_1(h) = 1$.

(d) There exists a number h', such that $f(h') > 0$ and $l_2(h') = 0$.

(e) $l_1(x) \leq 1$ for all x
$l_2(x) \geq 0$ for all x

(f) $\delta_1(x) + \delta_2(x) \geq \frac{1}{n}$ for all x, where n is the number of random, independent observations of the stochastic variable X.

Let $\varphi(x)$ be the distribution function of such a system of observations, i.e., the ratio, to n, of the number of observations $<x$ is $\phi(x)$. $\varphi(x)$ is, of course, a multiple of $\frac{1}{n}$ for all x.

We shall consider the following problem:

What is the probability P that

$$l_2(x) \leq \varphi(x) \leq l_1(x) \tag{1}$$

for all x?

The reasons for restrictions (b), (c), (d), (e), and (f) on $\delta_1(x)$ and $\delta_2(x)$ are now apparent. If there exist two numbers $q_1 < q_2$, such that, for $q_1 < x < q_2$, $l_1(x) > l_1(q_2)$ and $l_1(q_1) = l_1(q_2)$, then, if we change $l_1(x)$ so that $l_1(x) = l_1(q_2)$ for $q_1 \leq x \leq q_2$, P will remain unchanged. An analogous process leads to a similar conclusion for $l_2(x)$. Hence $l_1(x)$ and $l_2(x)$ are to be monotonically non-decreasing. If there did not exist a number h or h', P would be 0. Hence requirements (c) and (d). Since $0 \leq \varphi(x) \leq 1$, there is no point to considering functions which do not satisfy (e). $\varphi(x)$ is a step-function whose saltuses are $\geq \frac{1}{n}$. If, for all x.

$$\delta_1(x) + \delta_2(x) < \frac{1}{n}$$

then $P = 0$. If there is an interval $[\beta, \gamma]$ within which $\delta_1(x) + \delta_2(x) < \frac{1}{n}$, then all samples in which one of the observed values lies in this interval are

such that (1) does not hold for all x. For the sake of simplicity and because the situation described in (f) is the one of importance, we make the latter requirement.

It would appear that P depends upon $f(x)$, $\delta_1(x)$, $\delta_2(x)$, and n.

THEOREM: *P is independent of $f(x)$ and depends only upon $\delta_1(x)$, $\delta_2(x)$, and n.*

PROOF: Let $Y = f(X)$. Then Y is a stochastic variable distributed in the range 0 to 1 with a distribution function $\equiv x$. By this transformation $l_1(x)$ and $l_2(x)$ become respectively

$$\left.\begin{aligned} l_1'(x) &= x + \delta_1(x) \\ l_2'(x) &= x - \delta_2(x) \end{aligned}\right\} \quad 0 \leq x \leq 1. \tag{2}$$

Then P is the probability that the distribution function $\varphi(x)$ of a random sample of n of the stochastic variable Y shall be such that $l_2'(x) \leq \varphi(x) \leq l_1'(x)$ and is therefore independent of $f(x)$.

3. **Computation of P.** From the previous section it follows that, in computing P, we may confine ourselves to a stochastic variable X whose range is from 0 to 1 and whose distribution function $\equiv x$. Let $l_1(x)$ and $l_2(x)$ be the upper and lower limits, respectively, which are set for $\varphi(x)$. $l_1(x)$ and $l_2(x)$ are defined in (2), if the accents are omitted.

Consider the equations:

$$l_1(x) = \frac{i}{n} \qquad (i = 1, 2, \cdots, n; 0 \leq x \leq 1). \tag{3}$$

If, for a certain i, the corresponding equation possesses one or more solutions in x, let a_i be the minimum of these solutions. If the first r of these equations (3) have no solutions, let

$$a_i = 0 \qquad (i = 1, \cdots, r).$$

If the i^{th}, say, of the equations

$$l_2(x) = \frac{i-1}{n} \qquad (i = 1, \cdots, n; 0 \leq x \leq 1) \tag{4}$$

possesses one or more solutions in x, let b_i be the maximum of these. If the last $n - s$ of the equations (4) have no solutions, let

$$b_i = 1 \qquad (i = s + 1, \cdots, n).$$

Obviously

$$a_i \leq a_{i+1}, \qquad b_i \leq b_{i+1}, \qquad a_i \leq b_i.$$

From restrictions, (e) and (f) on $l_1(x)$ and $l_2(x)$, it follows that $a_1 = 0$, $b_n = 1$.

Suppose the sample $E = x_1, \cdots, x_n$ has been obtained. Arrange the x's

in ascending order, thus: $x_{p_1}, x_{p_2}, \cdots, x_{p_n}$ where $x_{p_1} \leq x_{p_2} \leq \cdots \leq x_{p_n}$. Then necessary and sufficient conditions that (1) hold are:

$$a_i \leq x_{p_i} \leq b_i \qquad (i = 1, \cdots, n). \tag{5}$$

Let $P_k(t, \Delta t)$, $(k = 0, 1, \cdots, (n-1); a_{k+1} \leq t \leq b_{k+1})$ be the probability that a sample $E = x_1, \cdots, x_n$ shall fulfill the following conditions:

(a) $x_1 \leq x_2 \leq \cdots \leq x_{k+1}$,

(b) $x_1, \cdots, x_k$ satisfy the first k inequalities (5),

(c) $t \leq x_{k+1} \leq t + \Delta t$.

Let

$$P_k(t) = \lim_{\Delta t \to 0} \frac{P_k(t, \Delta t)}{\Delta t}.$$

Since $f(x) \equiv x$, we get easily

$$P_0(t) \equiv 1. \tag{6}$$

We shall now develop a recursion formula for $P_{k+1}(t)$. For this purpose let us consider the following composite event: The observations $x_1, \cdots, x_n$ satisfy the conditions (a), (b), and

$$t' \leq x_{k+1} \leq t' + \Delta t'$$

and

$$t \leq x_{k+2} \leq t + \Delta t.$$

If $a_{k+1} \leq t' \leq b_{k+1}$, the probability of this event is $P_k(t', \Delta t')\Delta t$. Now

$$\lim_{\substack{\Delta t' \to 0 \\ \Delta t \to 0}} \frac{P_k(t', \Delta t')\Delta t}{\Delta t' \cdot \Delta t} = P_k(t').$$

$P_k(t')$ is obviously the probability density of the bivariate distribution of t' and t. In order to obtain $P_{k+1}(t)$ we have to integrate $P_k(t')\,dt'$ over the region defined by the two inequalities

$$t' \leq t$$

$$a_{k+1} \leq t' \leq b_{k+1}.$$

Hence, omitting the now unnecessary accent, if

$$t \leq b_{k+1} \tag{7}$$

then

$$P_{k+1}(t) = \int_{a_{k+1}}^{t} P_k(t)\,dt \qquad (k = 0, 1, \cdots, (n-2)), \tag{8}$$

and if

$$t > b_{k+1} \tag{9}$$

then

$$P_{k+1}(t) = \int_{a_{k+1}}^{b_{k+1}} P_k(t)\,dt \quad (k = 0, 1, 2, \cdots (n-2)). \tag{10}$$

Now, to obtain P, we cannot confine ourselves only to cases where $x_1 \leq x_2 \leq \cdots \leq x_n$, but have to consider all the $n!$ permutations of the n x's. Hence

$$P = n! \int_{a_n}^{b_n} P_{n-1}(t)\,dt. \tag{11}$$

The fact that there are two forms of the recursion formula corresponding to the two possible cases (7) and (9) makes actual calculation very cumbersome for n of any considerable size. We shall therefore give an approximation formula which is considerably easier to apply to practical calculations.

4. **Computation of $\overline{P}$ and $\underline{P}$.** Let $\overline{P}$ be the probability that, for a sample of n, $l_1(x) \geq \varphi(x)$ for all x. Let $\underline{P}$ be the probability that, for a sample of n, $\varphi(x) \geq l_2(x)$ for all x.

Consider the inequalities

$$\left.\begin{aligned} x_i &\geq a_i \quad (12)\\ x_i &\leq b_i \quad (13)\end{aligned}\right\} \qquad (i = 1, 2, \cdots, n)$$

Let

$$\overline{P}_k(t, \Delta t), \qquad (k = 0, 1, \cdots, (n-1); t \geq a_{k+1})$$

be the probability that a sample $E = x_1, \cdots, x_n$ of the stochastic variable X should fulfill the following conditions:

(a) $x_1 \leq x_2 \leq \cdots \leq x_{k+1}$

(b) $x_1, \cdots, x_k$ satisfy the first k inequalities (12)

(c) $t \leq x_{k+1} \leq t + \Delta t$.

Let

$$\overline{P}_k(t) = \lim_{\Delta t \to 0} \frac{\overline{P}_k(t, \Delta t)}{\Delta t}.$$

Then, by an argument like that employed in the preceding section, we obtain

$$\overline{P}_0(t) \equiv 1, \tag{14}$$

and the recursion formula

$$\overline{P}_{k+1}(t) = \int_{a_{k+1}}^{t} \overline{P}_k(t)\,dt. \tag{15}$$

Let $\overline{P}_n(t)$ be defined formally by (15). Then, in the same way in which we obtained (11), we get

$$\overline{P} = n!\,\overline{P}_n(1). \tag{16}$$

In the same manner we shall obtain an expression for $\underline{P}$.

Let $\underline{P}_k(t, \Delta t)$, $(k = 0, 1, \cdots (n-1); t \leq b_{n-k})$ be the probability that a sample $E = x_1, \cdots, x_n$ of the stochastic variable X should fulfill the following conditions:

(a) $x_{n-k} \leq x_{n-k+1} \leq \cdots \leq x_n$,

(b) $x_{n-k+1}, \cdots, x_n$ satisfy the last k inequalities (13),

(c) $t \leq x_{n-k} \leq t + \Delta t$.

Let

$$\underline{P}_k(t) = \lim_{\Delta t \to 0} \frac{\underline{P}_k(t, \Delta t)}{\Delta t}.$$

Then

$$\underline{P}_0(t) \equiv 1 \tag{17}$$

and by an argument very similar to that employed above,

$$\underline{P}_{k+1}(t) = \int_t^{b_{n-k}} \underline{P}_k(t)\, dt. \tag{18}$$

Let $\underline{P}_n(t)$ be defined formally by (18). Then

$$\underline{P} = n!\, \underline{P}_n(0). \tag{19}$$

The $\overline{P}_i(t)$ and $\underline{P}_i(t)$ are polynomials in t. Denote by c_j the constant term of $\overline{P}_j(t)$ and by d_j the constant term of $(-1)^j \underline{P}_j(t)$. Obviously

$$c_0 = 1 \tag{20}$$

$$d_0 = 1 \tag{21}$$

and

$$\overline{P}_i(t) = \frac{c_0}{i!} t^i + \frac{c_1}{(i-1)!} t^{i-1} + \cdots + c_{i-1} t + c_i \tag{22}$$

$$\underline{P}_i(t) = (-1)^i \left(\frac{d_0}{i!} t^i + \frac{d_1}{(i-1)!} t^{i-1} + \cdots + d_{i-1} t + d_i \right). \tag{23}$$

Since

$$\overline{P}_i(a_i) = 0, \qquad \underline{P}_i(b_{n-i+1}) = 0$$

we obtain

$$c_0 \frac{a_i^i}{i!} + c_1 \frac{a_i^{i-1}}{(i-1)!} + \cdots + c_{i-1} a_i + c_i = 0 \qquad (i = 1, 2, \cdots, n) \tag{24}$$

and

$$\frac{d_0}{i!} b_{n-i+1}^i + \frac{d_1}{(i-1)!} b_{n-i+1}^{i-1} + \cdots + d_{i-1} b_{n-i+1} + d_i = 0 \tag{25}$$

$$(i = 1, 2, \cdots, n)$$

The determinant of (20) and the first j equations (24) ($j = 1, \cdots, n$) considered as equations in $c_0, c_1, \cdots, c_j$ equals 1, since all the elements of the principal diagonal are 1 and all the elements above the principal diagonal are 0. Then

$$c_i = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 1 \\ a_1 & 1 & 0 & \cdots & 0 & 0 \\ \frac{a_2^2}{2!} & a_2 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{a_i^i}{i!} & \frac{a_i^{i-1}}{(i-1)!} & \frac{a_i^{i-2}}{(i-2)!} & \cdots & a_i & 0 \end{vmatrix} \tag{26}$$

$$= (-1)^i \begin{vmatrix} a_1 & 1 & 0 & \cdots & 0 \\ \frac{a_2^2}{2!} & a_2 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{a_i^i}{i!} & \frac{a_i^{i-1}}{(i-1)!} & \frac{a_i^{i-2}}{(i-2)!} & \cdots & a_i \end{vmatrix}.$$

From (16) and (22) for $i = n$, we get

$$\overline{P} = c_0 + nc_1 + n(n-1)c_2 + \cdots + n(n-1)\cdots(3)(2)c_{n-1} + n!\,c_n$$

$$= \begin{vmatrix} \frac{n!}{n!} & \frac{n!}{(n-1)!} & \frac{n!}{(n-2)!} & \cdots & \frac{n!}{1!} & \frac{n!}{0!} \\ a_1 & 1 & 0 & \cdots & 0 & 0 \\ \frac{a_2^2}{2!} & a_2 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{a_n^n}{n!} & \frac{a_n^{n-1}}{(n-1)!} & \frac{a_n^{n-2}}{(n-2)!} & \cdots & a_n & 1 \end{vmatrix}. \tag{27}$$

In the same way, we obtain

$$d_i = \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 & 1 \\ b_n & 1 & 0 & \cdots & 0 & 0 \\ \frac{b_{n-1}^2}{2!} & b_{n-1} & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{b_{n-i+1}^i}{i!} & \frac{b_{n-i+1}^{i-1}}{(i-1)!} & \frac{b_{n-i+1}^{i-2}}{(i-2)!} & \cdots & b_{n-i+1} & 0 \end{vmatrix} \tag{28}$$

$$= (-1)^i \begin{vmatrix} b_n & 1 & 0 & \cdots & 0 \\ \frac{b_{n-1}^2}{2!} & b_{n-1} & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{b_{n-i+1}^i}{i!} & \frac{b_{n-i+1}^{i-1}}{(i-1)!} & \frac{b_{n-i+1}^{i-2}}{(i-2)!} & \cdots & b_{n-i+1} \end{vmatrix}$$

and from (19) and (23) for $i = n$,

$$\underline{P} = (-1)^n n!\, d_n . \tag{29}$$

Perhaps if the determinants in (27) and (28) were to be simplified it might be easier to calculate $\overline{P}$ and $\underline{P}$ that way than by the recursion formulas.

5. **The approximation of P.** Let J be the probability that, for a sample of n, there exists at least one pair of numbers ω_1, ω_2, such that

$$0 \leq \omega_i \leq 1 \qquad (i = 1, 2)$$

$$\varphi(\omega_1) > l_1(\omega_1)$$
$$\varphi(\omega_2) < l_2(\omega_2).$$

Recalling the definitions of P, $\overline{P}$, and $\underline{P}$, it is obvious that

$$1 - P = (1 - \overline{P}) + (1 - \underline{P}) - J. \tag{30}$$

Now if

$$J \leq (1 - \overline{P})(1 - \underline{P}) \tag{31}$$

and $(1 - P)$ is small, the right member of (30) with J omitted furnishes an excellent approximation to $(1 - P)$. Suppose, for example, that it were desired to give upper and lower limits $l_1(x)$ and $l_2(x)$ such that $P = .95$. Choose $l_1(x)$ and $l_2(x)$ so that, for example, $\overline{P} = \underline{P} = .975$. Then P cannot differ from .95 by more than .000625. Even if

$$J \leq K(1 - \overline{P})(1 - \underline{P}) \tag{32}$$

where K is a small factor, say 10, the approximation would still be excellent. It seems very plausible that even (31) holds. However, we have not yet succeeded in obtaining a rigorous proof.

6. **The construction of confidence limits.** We now proceed to the construction of $l_{E,\alpha}(x)$ and $\underline{l}_{E,\alpha}(x)$ which were defined in Section I of this paper.

A confidence coefficient $\alpha(0 < \alpha < 1)$ is selected to which it is desired that the confidence limits correspond. Functions $\delta_1(x)$ and $\delta_2(x)$ are chosen to be as defined in Section 2 and also to be such as to make $P = \alpha$. This can be done by application of the formulas for the evaluation of P.

The functions $l_{E,\alpha}(x)$ and $\underline{l}_{E,\alpha}(x)$ are to be known when E and α are known.

Since α is given, $l_{E,\alpha}(x)$ and $\underline{l}_{E,\alpha}(x)$ depend upon the outcome of the experiment which yields observed values of the stochastic variable X. Let $E = x_1, \cdots, x_n$ be this system of values and let $\varphi(x)$ be its distribution function. Consider the equations

$$\delta_2(\varphi(x) + \Delta_1(x)) = \Delta_1(x) \tag{33}$$

$$\delta_1(\varphi(x) - \Delta_2(x)) = \Delta_2(x). \tag{34}$$

For a fixed but arbitrary x, $-\infty < x < +\infty$, $\varphi(x)$ is known and (33) and (34) are equations in $\Delta_1(x)$ and $\Delta_2(x)$. If, for a certain x, (33) has one or more solutions, let $\epsilon_1(x)$ be the maximum of the set of solutions (for this x, of course). Similarly, if for a certain x, (34) has one or more solutions, let $\epsilon_2(x)$ be the maximum of the set of solutions.

We can now give $l_{E,\alpha}(x)$ and $\underline{l}_{E,\alpha}(x)$ as follows:

For an x such that (33) has at least one solution,

$$l_{E,\alpha}(x) = \varphi(x) + \epsilon_1(x). \tag{35}$$

For an x such that (33) has no solutions,

$$l_{E,\alpha}(x) = 1. \tag{36}$$

For an x such that (34) has at least one solution,

$$\underline{l}_{E,\alpha}(x) = \varphi(x) - \epsilon_2(x). \tag{37}$$

For an x such that (34) has no solution,

$$\underline{l}_{E,\alpha}(x) = 0. \tag{38}$$

We recapitulate briefly the meaning of $l_{E,\alpha}(x)$ and $\underline{l}_{E,\alpha}(x)$ which were defined in Section 1. These are two functions defined for $-\infty < x < +\infty$ which may be constructed as above after a confidence coefficient α has been assigned and after the outcome of the physical experiment which determines the stochastic point E is known. These functions have the following property: No matter what the distribution function $f(x)$ of each of n stochastic independent variables $X_1, \cdots, X_n$ may be, provided only that $f(x)$ is continuous and the same for each $X_1, \cdots, X_n$, the probability is exactly α that, if we were to perform the physical experiment which gives a set of values E of the stochastic system $X_1, \cdots, X_n$ and were then to construct $l_{E,\alpha}(x)$ and $\underline{l}_{E,\alpha}(x)$, the inequality

$$\underline{l}_{E,\alpha}(x) \leq f(x) \leq l_{E,\alpha}(x) \tag{39}$$

would hold for all x.

A less precise but more intuitive statement of the above result is as follows: If, in many experiments we were to proceed as above to construct $l_{E,\alpha}(x)$ and $\underline{l}_{E,\alpha}(x)$ and then, in each instance, we were to predict that the unknown $f(x)$ (which need not be the same in all experiments) satisfies (39), the relative frequency of correct predictions would be α.

The formal proof of this result is exceedingly simple. For any continuous $f(x)$, the probability is α that

$$l_2(x) \leq \varphi(x) \leq l_1(x) \tag{40}$$

will hold for all x. This is so because of the way in which $\delta_1(x)$ and $\delta_2(x)$ were chosen. To prove the required result it would therefore be sufficient to show that, if (39) holds for all x, (40) holds for all x and conversely.

Let x be fixed but arbitrary. We shall show that

$$f(x) \leq l_{E,\alpha}(x) \tag{41}$$

implies

$$l_2(x) \leq \varphi(x) \tag{42}$$

and conversely.

If (33) has no solution, $\varphi(x) > l_2(1) \geq l_2(x)$, $l_{E,\alpha}(x) = 1$, and (41) and (42) are trivial. Assume therefore that (33) has at least one solution. For this situation, then, we have to show that

$$f(x) \leq \varphi(x) + \epsilon_1(x) \tag{43}$$

implies

$$l_2(x) \leq \varphi(x) \tag{44}$$

and conversely.

With x and hence $\varphi(x)$ and $\epsilon_1(x)$ fixed, consider the equation in x':

$$l_2(x') = \varphi(x). \tag{45}$$

Since $\varphi(x) \leq l_2(1)$, (45) has at least one solution. Let x'_m be the maximum of these solutions for a fixed x. Then from the definition of $\epsilon_1(x)$ it follows that

$$f(x'_m) - l_2(x'_m) = \epsilon_1(x), \tag{46}$$

or, on account of the definition of x'_m,

$$f(x'_m) = \varphi(x) + \epsilon_1(x). \tag{47}$$

Now, if (43) holds, $x \leq x'_m$ because of (47). Then, from the definition of x'_m and the fact that $l_2(x')$ is monotonically non-decreasing (44) follows.

If (44) holds, then $x \leq x'_m$ (by the definition of x'_m and the monotonic character of $l_2(x')$). Hence, because of (47), (43) is true. This shows the equivalence of (43) and (44).

In a similar manner, it may be shown that

$$l_{E,\alpha}(x) \leq f(x) \tag{48}$$

implies

$$\varphi(x) \leq l_1(x) \tag{49}$$

and conversely. This completes the proof.

7. **Miscellaneous remarks.** An expedient way of choosing $\delta_1(x)$ and $\delta_2(x)$ is such that, with c a constant,

$$\begin{aligned} x + \delta_1(x) &\equiv \min\,[x + c,\ 1] \\ x - \delta_2(x) &\equiv \max\,[x - c,\ 0]. \end{aligned} \qquad 0 \leq x \leq 1 \tag{50}$$

Tables of double entry could be constructed giving the c corresponding to specified α and n. With such tables available the construction of confidence limits would be quick and simple in practice. In this case, $\epsilon_1(x) = \epsilon_2(x) = c$.

Another expedient and plausible way of choosing $\delta_1(x)$ and $\delta_2(x)$ might be to choose them so that

$$\begin{aligned} x + \delta_1(x) &\equiv \min\,[px + q,\ 1] \\ x - \delta_2(x) &\equiv \max\,[p'x + q',\ 0] \end{aligned} \qquad 0 \leq x \leq 1 \tag{51}$$

where p, p', q, and q' are constants. The actual construction of confidence limits could then be handled with dispatch if similar tables were constructed.

$$l_{E,\alpha}(x) \quad \text{and} \quad \underline{l}_{E,\alpha}(x)$$

are, like $\varphi(x)$, step-functions. The situation may occur where, for $x = e$,

$$\lim_{(x<e),x\to e} l_{E,\alpha}(x) < \lim_{(x>e),x\to e} \underline{l}_{E,\alpha}(x).$$

This would give a prediction, corresponding to the confidence coefficient α, that $f(x)$ is not continuous. If $f(x)$ is continuous the probability of such a situation is 0.

8. **Further problems.** Even with α fixed, the functions $\delta_1(x)$ and $\delta_2(x)$ may be chosen in many ways. Each different choice gives, in general, different confidence limits. Which is to be preferred? This very problem also arose in the theory of parameter estimation and the testing of hypotheses and gave rise to the Neyman-Pearson theory. It would be desirable to develop such a theory for the confidence limits discussed in this paper.

We have treated here only the case where $f(x)$ is continuous. A similar theory is needed for the case where $f(x)$ is not continuous.

It would be of practical value to construct tables such as those described in Section 7. The construction of tables could be greatly facilitated if the formulas for P or $\overline{P}$ and $\underline{P}$ could be simplified so as to render them more practical for calculation or else if they were to be replaced by asymptotic expansions.

9. **An example.** To illustrate the method we shall consider an example for the case of samples of size 6, i.e. $n = 6$.

Let $\delta_1(x)$ and $\delta_2(x)$ be given as follows:

$$\delta_1(x) = d \quad \text{for} \quad 0 \leq x \leq 1 - d,$$

$$\delta_1(x) = 1 - x \quad \text{for} \quad 1 - d < x \leq 1,$$

$$\delta_2(x) = x \quad \text{for} \quad 0 \leq x \leq d,$$

and

$$\delta_2(x) = d \quad \text{for} \quad d < x \leq 1.$$

Denote by $\overline{P}$ the probability that

$$\varphi(x) \leq f(x) + \delta_1[f(x)],$$

by $\underline{P}$ the probability that

$$\varphi(x) \geq f(x) - \delta_2[f(x)]$$

and by P the probability that

$$f(x) - \delta_2[f(x)] \leq \varphi(x) \leq f(x) + \delta_1[f(x)].$$

$\varphi(x)$ denotes the sample distribution and $f(x)$ denotes the population distribution.

Since $\delta_2(x) = \delta_1(1 - x)$, we obviously have

$$\overline{P} = \underline{P}.$$

Let us calculate $\overline{P} = \underline{P}$ in case $d = \frac{1}{2}$. According to (3) we have

$$a_1 = a_2 = a_3 = 0, \qquad a_4 = \tfrac{1}{6}, \qquad a_5 = \tfrac{1}{3}, \qquad a_6 = \tfrac{1}{2}.$$

According to (16)

$$\overline{P} = 6!\overline{P}_6(1)$$

where

$$\overline{P}_0(t) \equiv 1,$$

$$\overline{P}_k(t) \equiv \int_{a_k}^{t} \overline{P}_{k-1}(t)\, dt \qquad (k = 1, \cdots, 6).$$

Applying this recursion formula we get

$$\overline{P}_1(t) = t; \qquad \overline{P}_2(t) = \frac{t^2}{2}, \qquad \overline{P}_3(t) = \frac{t^3}{6},$$

$$\overline{P}_4(t) = \frac{t^4}{24} - \frac{1}{2^7 \cdot 3^5}$$

$$\overline{P}_5(t) = \frac{t^5}{120} - \frac{t}{2^7 \cdot 3^5} - \frac{11}{3^6 \cdot 2^7 \cdot 5}$$

$$\overline{P}_6(t) = \frac{t^6}{720} - \frac{t^2}{2^8 \cdot 3^5} - \frac{11t}{3^6 \cdot 2^7 \cdot 5} - \frac{11}{2^9 \cdot 3^6 \cdot 5}.$$

Hence

$$\overline{P} = \underline{P} = 6!\overline{P}_6(1) = 1 - \frac{85}{2592} = 0.967.$$

Let us now calculate $\overline{P} = \underline{P}$ in case $d = \frac{1}{3}$. We have

$$a_1 = a_2 = 0, \qquad a_3 = \tfrac{1}{6}, \qquad a_4 = \tfrac{1}{3}, \qquad a_5 = \tfrac{1}{2} \quad \text{and} \quad a_6 = \tfrac{2}{3}.$$

Applying the recursion formula we get

$$\overline{P}_0(t) = 1, \qquad \overline{P}_1(t) = t, \qquad \overline{P}_2(t) = \frac{t^2}{2}, \qquad \overline{P}_3(t) = \frac{t^3}{6} - \frac{1}{2^4 \cdot 3^4},$$

$$\overline{P}_4(t) = \frac{t^4}{24} - \frac{t}{2^4 \cdot 3^4} - \frac{1}{2^4 \cdot 3^5},$$

$$\overline{P}_5(t) = \frac{t^5}{120} - \frac{t^2}{2^5 \cdot 3^4} - \frac{t}{2^4 \cdot 3^5} - \frac{11}{2^8 \cdot 3^5 \cdot 5},$$

$$\overline{P}_6(t) = \frac{t^6}{720} - \frac{t^3}{2^5 \cdot 3^5} - \frac{t^2}{2^5 \cdot 3^5} - \frac{11t}{2^8 \cdot 3^5 \cdot 5} - \frac{13}{2^7 \cdot 3^8 \cdot 5}.$$

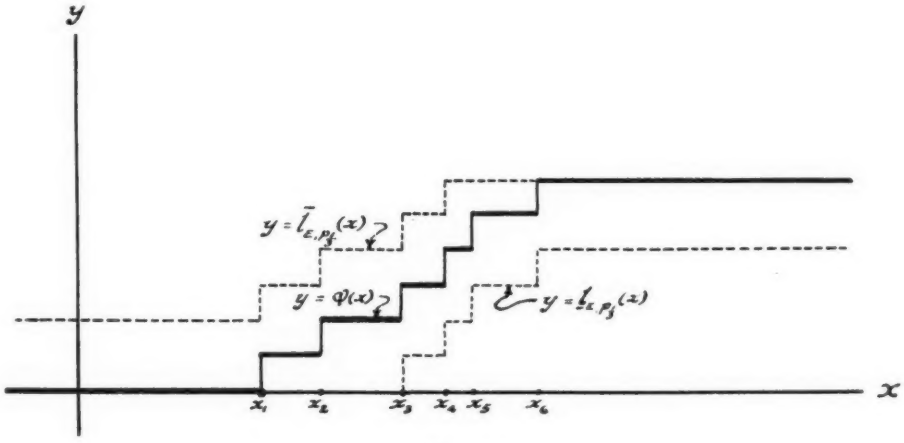

FIG. 1

Hence

$$\overline{P} = \underline{P} = 6!\overline{P}_6(1) = 1 - \frac{2483}{11664} = 0.787.$$

It is obvious that

$$1 - P = (1 - \overline{P}) + (1 - \underline{P}) - J,$$

where J denotes the probability that $\varphi(x)$ violates both limits. In case $d = \frac{1}{2}$ no $\varphi(x)$ exists which violates both limits, and therefore $J = 0$. If $d = \frac{1}{3}$,

J is not zero but so small that it can be neglected. Hence

$$P = 0.934 \quad \text{if} \quad d = \tfrac{1}{2}$$

and

$$P = 0.574 \quad \text{if} \quad d = \tfrac{1}{3}$$

P increases monotonically from 0.574 to 0.934 if d increases from $\frac{1}{3}$ to $\frac{1}{2}$. Denote by P_d the probability corresponding to d. According to (33)–(38), the confidence limits corresponding to the probability level P_d are given as follows:

$$l_{E,P_d}(x) = \varphi(x) + d \text{ if } \varphi(x) + d \leq 1,$$

$$l_{E,P_d}(x) = 1 \text{ if } \varphi(x) + d > 1,$$

$$l_{E,P_d}(x) = \varphi(x) - d \text{ if } \varphi(x) - d \geq 0$$

and

$$l_{E,P_d}(x) = 0 \text{ if } \varphi(x) - d < 0.$$

Substituting for d the numbers $\frac{1}{2}$ and $\frac{1}{3}$, we get the confidence limits corresponding to the probability levels 0.934 and 0.574 respectively. The upper and lower confidence limits for the *population* distribution corresponding to the probability level 0.574 are represented geometrically in Figure 1 by the upper and lower dotted broken lines for a sample of 6 having the values $x_1, x_2, \cdots x_6$. The sample distribution is represented by the solid broken line.

COLUMBIA UNIVERSITY
AND
BROOKLYN, N. Y.

ON THE POWER OF THE L_1 TEST FOR EQUALITY OF SEVERAL VARIANCES

By George W. Brown

The criterion L_1 was obtained by Neyman and Pearson[1] for testing the statistical hypothesis H_1 that k samples, known to be from normal universes, are actually from universes with equal variances, where the means are unspecified. The test seems to be of importance, when one considers the number of experiments which are concerned with the comparison of several types of treatments. The experimenter is in most cases interested in the respective means, and it is usually assumed, in order to test for significance of the difference between sample means, that the variances of the distributions are equal. At present, significance tests for justifying this assumption are rarely applied. Because of the unsatisfactory status of the problem of testing simultaneously for means and variances, the L_1 test is appropriate for justifying first the assumption of equal variances before testing for the means.

Neyman and Pearson have treated the sampling distribution of L_1 when H_1 is true, and Wilks and Thompson[2] have discussed the general distribution of the criterion when H_1 is not true. Here we shall show that the test is unbiassed when the number of observations is the same in each sample, and is in general unbiassed in the limit, in a certain sense. In addition, values of the power function have been computed for a few selected cases, when k is 2, in order to exhibit qualitatively the sharpness of the test.

Let the i-th sample ($i = 1, 2, \cdots, k$) of n_i individuals be denoted by Σ_i and suppose Σ_i has been drawn at random from a normal population with mean m_i (unknown) and variance $\sigma_i^2 = \frac{1}{A_i}$. Denote the observations of Σ_i by x_{ir} ($r = 1, 2, \cdots, n_i$). Then the criterion L_1 is expressible[3] in terms of the observations as follows:

$$L_1^{\frac{1}{2}n} = \frac{n^{\frac{1}{2}n} \prod_{i=1}^{k} (c_i^2)^{\frac{1}{2}n_i}}{\prod_{i=1}^{k} n_i^{\frac{1}{2}n_i} \left[\sum_{i=1}^{k} c_i^2\right]^{\frac{1}{2}n}} \tag{1}$$

where $n = \Sigma n_i$ and $c_i^2 = \sum_{r=1}^{n_i} (x_{ir} - \bar{x}_i)^2$. For convenience we shall let $L_1^{\frac{1}{2}n} = \lambda$.

[1] [1], pp. 461–464.

[2] See [4]. Nayer [3], studied the Type I approximation to the criterion L_1 and tabulated significance limits, etc.

[3] See [1], p. 464.

The variables $A_i c_i^2$ are independently distributed according to χ^2-laws with $n_i - 1$ degrees of freedom, respectively, hence the joint distribution of the c_i^2, when $\frac{1}{A_i}$ is the true value of σ_i^2 $(i = 1, 2, \cdots, k)$, is given by

$$(2) \qquad \frac{1}{2^{\frac{1}{2}n} \prod_i \Gamma\left(\frac{n_i - 1}{2}\right)} \cdot \prod_i [A_i^{\frac{1}{2}(n_i-1)} (c_i^2)^{\frac{1}{2}(n_i-3)}] e^{-\frac{1}{2}\sum A_i c_i^2} \, dc_1^2 \cdots dc_k^2$$

The power function,[4] which is defined as the probability of rejecting H_1, is given by $P(\lambda < \lambda_0)$, and is a function of the true values of the parameters $A_1, \cdots, A_k$, where λ_0 is defined so that $P(\lambda < \lambda_0) = \alpha$ when H_1 is true. Thus

$$F(A_1, \cdots, A_k) = P(\lambda < \lambda_0)$$

$$(3) \qquad = \frac{1}{2^{\frac{1}{2}n} \prod_i \Gamma\left(\frac{n_i - 1}{2}\right)} \int_{\lambda<\lambda_0} \prod_i [A_i^{\frac{1}{2}(n_i-1)} (c_i^2)^{\frac{1}{2}(n_i-3)}] e^{-\frac{1}{2}\sum A_i c_i^2} \, dc_1^2 \cdots dc_k^2$$

Note that when H_1 is true $P(\lambda < \lambda_0)$ is independent of the actual common value of the parameters, because of the homogeneity of λ.

Let us now restrict ourselves to the case in which $n_i = p$, $n = kp$. (1) and (3) become

$$(1') \qquad \lambda = k^{\frac{1}{2}kp} \left\{ \frac{\Pi c_i^2}{[\Sigma c_i^2]^k} \right\}^{\frac{1}{2}p}$$

and

$$F(A_1, A_2, \cdots, A_k)$$

$$(3') \qquad = \frac{1}{\left[2^{\frac{1}{2}p} \Gamma\left(\frac{p-1}{2}\right)\right]^k} \int_{\lambda<\lambda_0} \prod_{i=1}^{k} A_i^{\frac{1}{2}(p-1)} (c_i^2)^{\frac{1}{2}(p-3)} e^{-\frac{1}{2}\sum A_i c_i^2} \, dc_1^2 \cdots dc_k^2$$

We shall prove the following

THEOREM: *If* $n_1 = n_2 = \cdots = n_k = p$, *then* $F(A_1, A_2, \cdots, A_k) \geq F(A, A, \cdots, A)$. In other words, the probability of rejecting H_1 when H_1 is true is less than or at most equal to the probability of rejecting the hypothesis when any alternative is true, that is, the test is unbiassed. It should be noted that the statement of the theorem is to hold for each value of λ_0.

It is evident that $F(A_1, A_2, \cdots, A_k)$ remains invariant under permutations of the arguments, because of the symmetry in the c_i^2 of λ and of the integrand in (3'). Moreover, by using the homogeneity of λ we obtain the following relations

$$(4) \quad F(A_1, A_2, \cdots, A_k) = F\left(\frac{A_1}{A_k}, \frac{A_2}{A_k}, \cdots, \frac{A_{k-1}}{A_k}, 1\right) = F\left(1, \frac{A_2}{A_1}, \cdots, \frac{A_{k-1}}{A_1}, \frac{A_k}{A_1}\right)$$

[4] Defined by Neyman and Pearson, [2], p. 5.

Now if we set $a_i = \frac{A_i}{A_k}$ $(i = 1, 2, \cdots, k-1)$, we may replace $F(A_1, A_2, \cdots, A_k)$ by $F(a_1, a_2, \cdots, a_{k-1}, 1) = f(a_1, \cdots, a_k)$; we must now show that $f(a_1, \cdots, a_{k-1}) \geq f(1, 1, \cdots, 1)$. From (4) we obtain

$$F(a_1, a_2, \cdots, a_{k-1}, 1) = F\left(1, \frac{a_2}{a_1}, \frac{a_3}{a_1}, \cdots, \frac{a_{k-1}}{a_1}, \frac{1}{a_1}\right) \tag{4'}$$

and permuting the arguments we have, finally,

$$f(a_1, a_2, \cdots, a_{k-1}) = f\left(\frac{1}{a_1}, \frac{a_2}{a_1}, \frac{a_3}{a_1}, \cdots, \frac{a_{k-1}}{a_1}\right). \tag{5}$$

Differentiate (5) with respect to a_1,

$$\begin{aligned} f_1(a_1, a_2, \cdots, a_{k-1}) = -\frac{1}{a_1^2}\Big[& f_1\left(\frac{1}{a_1}, \frac{a_2}{a_1}, \cdots, \frac{a_{k-1}}{a_1}\right) \\ & + a_2 f_2\left(\frac{1}{a_1}, \frac{a_2}{a_1}, \cdots\right) + \cdots + a_{k-1} f_{k-1}\left(\frac{1}{a_1}, \frac{a_2}{a_1}, \cdots\right)\Big] \end{aligned} \tag{6}$$

and set $a_1 = a_2 = \cdots = a_{k-1} = 1$, obtaining $f_1(1, 1, \cdots, 1) = -\sum_{i=1}^{k-1} f_i(1, 1, \cdots, 1)$. But $f_i(1, 1, \cdots, 1) = f_j(1, 1, \cdots, 1)$, hence

$$f_i(1, 1, \cdots, 1) = 0; \qquad i = 1, 2, \cdots, k-1. \tag{7}$$

Now differentiating (6) with respect to a_1 and evaluating at $a_i = 1$, we have $f_{11}(1, 1, \cdots, 1) = \sum_{ij} f_{ij}(1, 1, \cdots, 1)$, that is,

$$\begin{aligned} f_{11}(1, 1, \cdots, 1) - f_{11}(1, 1, \cdots, 1) - f_{22}(1, 1, \cdots, 1) \\ - \cdots - f_{k-1\,k-1}(1, 1, \cdots, 1) = \sum_{i \neq j} f_{ij}(1, 1, \cdots, 1), \end{aligned}$$

hence, by the symmetry of the variables,

$$\begin{aligned} f_{ij}(1, 1, \cdots, 1) &= -\frac{1}{k-1} f_{11}(1, 1, \cdots, 1), \qquad i \neq j; \\ f_{ii}(1, 1, \cdots, 1) &= f_{11}(1, 1, \cdots, 1). \end{aligned} \tag{8}$$

It is easily verified from (8) that the $f_{ij}(1, 1, \cdots, 1)$ are coefficients of a definite quadratic form in $k-1$ variables. Therefore there is an extremum at $(1, 1, \cdots, 1)$. It remains to show that $f_{11}(1, 1, \cdots, 1) > 0$ in order to establish that the extremum is actually a minimum.

In (3') we make the transformation $u_i = A_k \frac{c_i^2}{c_k^2}$; $i = 1, \cdots, k-1$; $u_k = A_k c_k^2$, and integrate out the variable u_k, since λ is now independent of u_k, obtaining

$$(9)\quad f(a_1, a_2, \cdots, a_k) = B\prod_{i=1}^{k-1} a_i^{\frac{1}{2}(p-1)} \int_{\lambda<\lambda_0} \frac{\prod_{i=1}^{k-1} u_i^{\frac{1}{2}(p-3)}}{\left[1 + \sum_{i=1}^{k-1} a_i u_i\right]^{\frac{1}{2}k(p-1)}}\, du_1 \cdots du_{k-1}$$

$$(10)\qquad \lambda = k^{\frac{1}{2}kp}\left\{\frac{\prod_{i=1}^{k-1} u_i}{\left[1 + \sum_{i=1}^{k-1} u_i\right]^k}\right\}^{\frac{1}{2}p}; \qquad u_i > 0$$

where B is some positive constant independent of the a_i. From (9)

$$f_1 = B\prod_{i=1}^{k-1} a_i^{\frac{1}{2}(p-1)} \int_{\lambda<\lambda_0} \left\{\frac{p-1}{2a_1} - \frac{k(p-1)u_1}{2\left[1 + \sum_{i=1}^{k-1} a_i u_i\right]}\right\}$$

$$(11)\qquad \frac{\prod_{i=1}^{k-1} u_i^{\frac{1}{2}(p-3)}}{\left[1 + \sum_{i=1}^{k-1} a_i u_i\right]^{\frac{1}{2}k(p-1)}}\, du_1 \cdots du_{k-1}$$

The last step involves differentiation under the sign of integration, which is certainly justifiable here.

Now consider λ for fixed $u_2, \cdots, u_{k-1}$, and variable u_1. $\lambda < \lambda_0$ is equivalent to the statement $\frac{u_1}{[\varphi + u_1]^k} < \theta$ where φ and θ depend on $u_2, u_3, \cdots, u_{k-1}$; $\varphi, \theta > 0$. The function $\psi(u_1) = \frac{u_1}{(\varphi + u_1)^k}$ has a maximum at $u_1 = \frac{\varphi}{k-1}$, and has no other extrema, hence the equation $\frac{u_1}{(\varphi + u_1)^k} = \theta$ has but two positive roots, x_1 and x_2, say. Let $x_2 > x_1$. Then for fixed $u_2, u_3, \cdots, u_{k-1}$ the region $\lambda < \lambda_0$ is composed of the u_1 intervals $(0, x_1)$ and (x_2, ∞). Now examining the integrand in (11) we see that it is the partial derivative with respect to u_1 of the quantity

$$\frac{1}{a_1}\frac{u_1^{\frac{1}{2}(p-1)} \prod_{2}^{k-1} u_i^{\frac{1}{2}(p-3)}}{\left[1 + \sum_{1}^{k-1} a_i u_i\right]^{\frac{1}{2}k(p-1)}}.$$

This quantity vanishes at 0 and ∞, hence

$$(12)\quad f_1 = \frac{1}{a_1} B \cdot \prod_{i=1}^{k-1} a_i^{\frac{1}{2}(p-1)} \int_G \prod_{2}^{k-1} u_i^{\frac{1}{2}(p-3)} \left[\frac{u_1^{\frac{1}{2}(p-1)}}{\left(1 + \sum_{1}^{k-1} a_i u_i\right)^{\frac{1}{2}k(p-1)}}\right]_{x_2}^{x_1} du_2 \cdots du_{k-1}$$

where G is some region of positive measure in the space of the variables

$u_2, u_3, \cdots, u_{k-1}$. Now differentiating in (12) with respect to a_1, and setting $a_1 = a_2 = \cdots = a_{k-1} = 1$, we get

$$f_{11}(1, 1, \cdots, 1) = B \int_G \prod_2^{k-1} u_i^{\frac{1}{2}(p-3)} \left\{ \frac{p-3}{2} \left[\frac{u_1^{\frac{1}{2}(p-1)}}{(\varphi + u_1)^{\frac{1}{2}k(p-1)}} \right]_{x_2}^{x_1} - \left[\frac{k(p-1)u_1^{\frac{1}{2}(p+1)}}{2(\varphi + u_1)^{\frac{1}{2}k(p-1)+1}} \right]_{x_2}^{x_1} \right\} du_2 \cdots du_{k-1}$$

The first term inside the braces has the value $\theta^{\frac{1}{2}(p-1)}$ both at x_1 and x_2, hence vanishes when evaluated between those limits, so that

$$f_{11}(1, 1, \cdots, 1) = \frac{k(p-1)}{2} B \int_G \prod_2^{k-1} u_i^{\frac{1}{2}(p-3)} \left\{ \frac{x_2^{\frac{1}{2}(p+1)}}{(\varphi + x_2)^{\frac{1}{2}k(p-1)+1}} - \frac{x_1^{\frac{1}{2}(p+1)}}{(\varphi + x_1)^{\frac{1}{2}k(p-1)+1}} \right\} du_2 \cdots du_{k-1} \tag{13}$$

x_1 and x_2 are roots of the equation $\frac{u_1}{(\varphi + u_1)^k} = \theta$, hence $x_1 = \theta(\varphi + x_1)^k$ and $x_2 = \theta(\varphi + x_2)^k$. Putting these values in the numerators in (13), we have

$$f_{11}(1, 1, \cdots, 1) = \frac{k(p-1)}{2} B \int_G \theta^{\frac{1}{2}(p+1)} \prod_2^{k-1} u_i^{\frac{1}{2}(p-3)} \{(\varphi + x_2)^{k-1} - (\varphi + x_1)^{k-1}\} du_2 \cdots du_{k-1}.$$

The integrand is positive, since $\theta, \varphi > 0$ and $x_2 > x_1$, hence $f_{11}(1, 1, \cdots, 1) > 0$. We have shown, then, that the power function has a relative minimum, at least, when H_1 is true. We shall show that the minimum is in fact an absolute minimum.

Consider the integrand in (12). The integrand has the same sign as

$$\frac{x_1^{\frac{1}{2}(p-1)}}{\left(1 + a_1 x_1 + \sum_2^{k-1} a_i u_i\right)^{\frac{1}{2}k(p-1)}} - \frac{x_2^{\frac{1}{2}(p-1)}}{\left(1 + a_1 x_2 + \sum_2^{k-1} a_i u_i\right)^{\frac{1}{2}k(p-1)}}.$$

But $x_1 = \theta(1 + x_1 + \Sigma u_i)^k$ and $x_2 = \theta(1 + x_2 + \Sigma u_i)^k$. Hence the integrand has the same sign as

$$\frac{1 + x_1 + \sum_2^{k-1} u_i}{1 + a_1 x_1 + \sum_2^{k-1} a_i u_i} - \frac{1 + x_2 + \sum_2^{k-1} u_i}{1 + a_1 x_2 + \sum_2^{k-1} a_i u_i},$$

so that the integrand is positive or negative accordingly, as $(x_1 - x_2)\left[1 + \sum_2^{k-1} a_i u_i - a_1\left(1 + \sum_2^{k-1} u_i\right)\right]$ is positive or negative. Since $x_1 < x_2$, this last quantity is positive if $a_1 > 1$ and $a_i \leq a_1$, and negative if $a_1 < 1$ and $a_i \geq a_1$. Hence we conclude that $\frac{\partial f}{\partial a_1} > 0$ if $a_1 > 1$ and $a_i \leq a_1$, and $\frac{\partial f}{\partial a_1} < 0$

if $a_1 < 1$ and $a_i \geq a_1$. By the symmetry in the variables the same is true of $\frac{\partial f}{\partial a_i}$, i.e., $\frac{\partial f}{\partial a_i} > 0$ if $a_i > 1$ and $a_i = \max(a_j)$, and $\frac{\partial f}{\partial a_i} < 0$ if $a_i < 1$ and $a_i = \min(a_j)$. Now suppose $(a_1^0, \cdots, a_k^0) \neq (1, \cdots, 1)$. Then either $\max(a_i^0) > 1$ or $\min(a_i^0) < 1$. Hence the first partials can vanish simultaneously only at $(1, 1, \cdots, 1)$, so that f can have no other extrema. Therefore f must have an absolute minimum at $(1, 1, \cdots, 1)$. This completes the proof that the L_1 test is unbiassed when $n_1 = n_2 = \cdots = n_k$.

It is easily seen that the test is in general biassed when the samples consist of different numbers of observations. Consider the case $k = 2$, with samples of n_1 and n_2 observations respectively. In this case we have the single parameter $a = \frac{A_1}{A_2}$. As in (9) and (10),

$$f(a) = Ba^{\frac{1}{2}(n_1-1)} \int_{\lambda<\lambda_0} \frac{u^{\frac{1}{2}(n_1-3)}}{(1+au)^{\frac{1}{2}n-1}}\, du \tag{14}$$

$$\lambda = \left(\frac{n^{\frac{1}{2}n}}{n_1^{\frac{1}{2}n_1} n_2^{\frac{1}{2}n_2}}\right) \frac{u^{\frac{1}{2}n_1}}{(1+u)^{\frac{1}{2}n}}. \tag{15}$$

As before, the equation $\lambda = \lambda_0$ has but two positive roots, $x_2 > x_1 > 0$, so that, as in (12),

$$f'(a) = Ba^{\frac{1}{2}(n_1-3)}\left[\frac{u^{\frac{1}{2}(n_1-1)}}{(1+au)^{\frac{1}{2}n-1}}\right]_{x_2}^{x_1}$$

$$= Ba^{\frac{1}{2}(n_1-3)}\left[\frac{x_1^{\frac{1}{2}(n_1-1)}}{(1+ax_1)^{\frac{1}{2}n-1}} - \frac{x_2^{\frac{1}{2}(n_1-1)}}{(1+ax_2)^{\frac{1}{2}n-1}}\right].$$

Therefore $f'(1) = B\left[\frac{x_1^{\frac{1}{2}(n_1-1)}}{(1+x_1)^{\frac{1}{2}n-1}} - \frac{x_2^{\frac{1}{2}(n_1-1)}}{(1+x_2)^{\frac{1}{2}n-1}}\right]$.

Recalling that $\frac{x_1^{n_1}}{(1+x_1)^n} = \frac{x_2^{n_1}}{(1+x_2)^n}$ it is evident that $f'(1) = 0$ if and only if $n_1 = \frac{n}{2}$. Hence if $n_1 \neq n_2$, the power function does not have a minimum at $a = 1$. It can be shown in this case that a minimum does exist at some point, and if $n \to \infty$ so that $n_1 = \alpha_1 n$, then the minimum tends to the point $a = 1$. The proof is omitted, in view of the fact that a general result of a different nature will be obtained.

Before proceeding, we shall establish a lemma which is undoubtedly well known. However, on account of the directness of the argument, the proof is given here.

LEMMA: *If $x_1, x_2, \cdots, x_k$ have joint distribution function $f_n(x_1, x_2, \cdots, x_k)$ such that $E(x_i) \underset{n\to\infty}{\to} m_i$ and $E[(x_i - E(x_i))^2] \underset{n\to\infty}{\to} 0$, and if $y = \varphi(x_1, x_2, \cdots, x_k)$ is continuous in $x_1, x_2, \cdots, x_k$ at the point $(m_1, m_2, \cdots, m_k)$, then the distribution of y converges stochastically to the point $\varphi(m_1, m_2, \cdots, m_k)$.*

Proof: By Tschebycheff's Inequality,

$$P\left\{|x_i - E(x_i)| > \frac{\delta}{2}\right\} \leq \frac{4}{\delta^2} E[(x_i - E(x_i))^2].$$

Let n be large enough so that $|E(x_i) - m_i| < \frac{\delta}{2}$; $i = 1, 2, \cdots, h$. Then $|x_i - m_i| > \delta$ implies $|x_i - E(x_i)| > \frac{\delta}{2}$, hence

$$P\{|x_i - m_i| > \delta\} \leq \frac{4}{\delta^2} E[(x_i - E(x_i))^2].$$

Let w_δ denote a cube a side 2δ about the point $(m_1, \cdots, m_h)$, and let x denote the point $(x_1, \cdots, x_h)$.

$$P[x \notin w_\delta] \leq \sum_{i=1}^{h} P\{|x_i - m_i| > \delta\},$$

hence

$$P[x \notin w_\delta] \leq \frac{4}{\delta^2} \sum_{i=1}^{h} E[(x_i - E(x_i))^2],$$

therefore $P[x \notin w_\delta] \underset{n\to\infty}{\to} 0$, that is $P[x \subset w_\delta] \to 1$. Given any interval w'_ϵ about the point $y = \varphi(m_1, m_2, \cdots, m_h)$, there is a cube w_δ about $(m_1, m_2, \cdots, m_h)$ such that $x \subset w_\delta$ implies $y \subset w'_\epsilon$. $P[x \subset w_\delta] \leq P[y \subset w'_\epsilon]$, but $P[x \subset w_\delta] \to 1$, therefore $P[y \subset w'_\epsilon] \to 1$. That is, y converges stochastically to the point $y = \varphi(m_1, m_2, \cdots, m_h)$.

Referring to (1), we may express λ as a function of $k - 1$ variables as follows:

$$\lambda = \frac{n^{\frac{1}{2}n} \prod_{i=1}^{k-1} u_i^{\frac{1}{2}n_i}}{\prod_{i=1}^{k} n_i^{\frac{1}{2}n_i} \left[1 + \sum_{i=1}^{k-1} u_i\right]^{\frac{1}{2}n}}$$

where $u_i = \frac{c_i^2}{c_k^2}$; $i = 1, 2, \cdots, k - 1$. Let $n \to \infty$, and let $n_i = \alpha_i n$, $\Sigma\alpha_i = 1$. Then

$$\lambda^{\frac{2}{n}} = \frac{\prod_{i=1}^{k-1} u_i^{\alpha_i}}{\prod_{i=1}^{k} \alpha_i^{\alpha_i} \left[1 + \sum_{i=1}^{k-1} u_i\right]}.$$

From (2) it is seen that $E(u_i) = E\left(\frac{c_i^2}{c_k^2}\right) = \frac{A_k}{A_i}\frac{n_i - 1}{n_k - 1} = \frac{1}{a_i} \cdot \frac{n_i - 1}{n_k - 1}$, and $E(u_i^2) = \left(\frac{1}{a_i}\right)^2 \frac{(n_i - 1)(n_i + 1)}{(n_k - 1)(n_k + 1)}$. Therefore $E(u_i) \to \frac{1}{a_i}\frac{\alpha_i}{\alpha_k}$ and $E(u_i^2) \to \left(\frac{1}{a_i}\frac{\alpha_i}{\alpha_n}\right)^2$ in other

words, $E[(u_i - E(u_i))^2] \to 0$. Now we apply the lemma, concluding that λ_n^2, that is, L_1 converges stochastically to the quantity

$$r = \frac{1}{\prod_{i=1}^{k-1} a_i^{\alpha_i} \left[\alpha_k + \sum_{i=1}^{k-1} \frac{\alpha_i}{a_i}\right]}.$$

TABLE I

(1) $n_1 = 5, n_2 = 5$		(2) $n_1 = 10, n_2 = 10$		(3) $n_1 = 20, n_2 = 20$	
a	$f(a)$	a	$f(a)$	a	$f(a)$
1	.05	1	.05	1	.05
4/3	.06	4/3	.07	4/3	.09
2	.09	2	.16	2	.31
3	.15	3	.33	3	.65
4	.21	4	.50	4	.84
5	.27	5	.62	5	.93
10	.52	6	.72		
20	.75	10	.90		

TABLE II

(1) $n_1 = 12, n_2 = 8$		(2) $n_1 = 15, n_2 = 10$	
a	$f(a)$	a	$f(a)$
1/10	.90	1/10	.96
1/5	.61	1/5	.74
1/4	.47	1/4	.59
1/3	.32	1/3	.40
1/2	.16	1/2	.20
3/5	.11	3/5	.11
4/5	.07	1	.05
1	.05	1.5	.09
1.2	.05	2	.17
1.5	.06	3	.38
2	.13	4	.60
3	.30	5	.70
4	.45	10	.95
5	.60		
6	.67		
10	.87		

r is the ratio of weighted geometric mean to arithmetic mean of the quantities $1, \frac{1}{a_1}, \frac{1}{a_2}, \cdots, \frac{1}{a_{k-1}}$, hence $r = 1$ if and only if $a_1 = a_2 = \cdots = a_{k-1} = 1$, otherwise $r < 1$. Therefore when H_1 is true $\lambda^{\frac{2}{n}}$ converges stochastically to 1, otherwise $\lambda^{\frac{2}{n}}$ converges stochastically to some value less than 1.

Let us choose $\lambda_0^{(n)}$ so that $P(\lambda < \lambda_0^{(n)}) = \alpha$ when H_1 is true. Consider some alternative hypothesis H_1^*. $\lambda^{\frac{2}{n}}$ converges stochastically to $r < 1$. Choose ζ so that $r < \zeta < 1$. $P(\lambda^{\frac{2}{n}} < \zeta) \to 0$ when H_1 is true, but $P(\lambda^{\frac{2}{n}} < \lambda_0^{(n)\frac{2}{n}}) = \alpha$

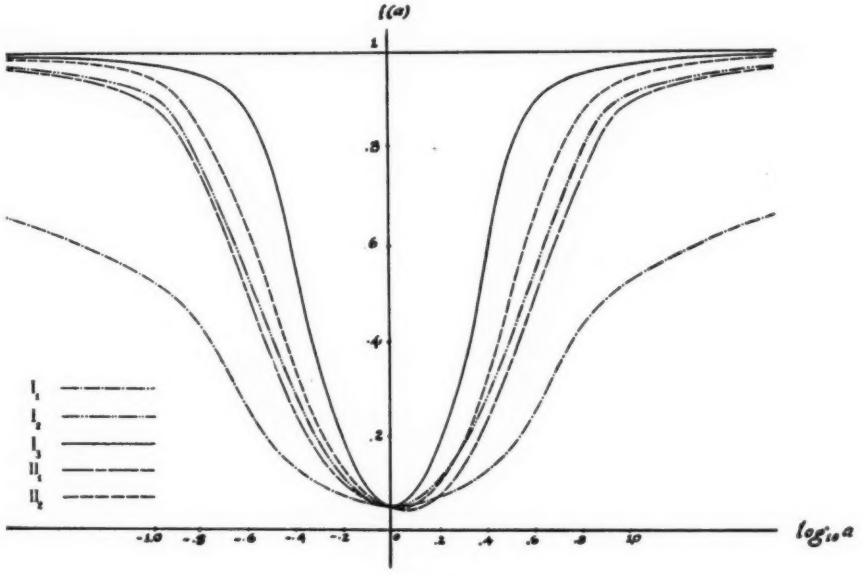

FIG. 1

when H_1 is true, thus, for n sufficiently large, $\zeta < \lambda_0^{(n)\frac{2}{n}}$, that is, $\zeta^{\frac{n}{2}} < \lambda_0^{(n)}$. Therefore $P(\lambda < \lambda_0^{(n)}) \geq P(\lambda < \zeta^{\frac{n}{2}}) = P(\lambda^{\frac{2}{n}} < \zeta)$. Now, if H_1^* is true, $P(\lambda^{\frac{2}{n}} < \zeta) \to 1$, therefore $P(\lambda < \lambda_0^{(n)}) \to 1$.

We have shown then, that *if $n \to \infty$ so that $n_i = \alpha_i n$, where the α_i are fixed, while the probability level α remains constant, then the power of the test with respect to any alternative hypothesis H_1^* tends to unity.* It is impossible, of course, to have the power function tend to unity uniformly with respect to all alternative hypotheses, since the power function is continuous for all n, and since the power with respect to H_1 is constantly α. What we can conclude, however, is that for any particular alternative hypothesis, the probability of rejecting H_1

is greater than α for sufficiently large n.[5] (We might say, then, that the test is asymptotically unbiassed.) Moreover, the fact that the power with respect to H_1^* tends to unity implies that the test becomes sharper with increasing n.

In order to illustrate the sharpness of the test, values of the power function were computed, when $k = 2$, for the cases $n_1 = n_2 = 5$; $n_1 = n_2 = 10$; $n_1 = n_2 = 20$; $n_1 = 12$, $n_2 = 8$; and $n_1 = 15$, $n_2 = 10$. The results are given in Tables I and II. The computations were made from (14) and (15) by means of Pearson's *Tables of the Incomplete Beta Function*. The roots x_1 and x_2 of the equation $\lambda = \lambda_0$ were determined, for $\alpha = .05$, by trial and error, making it possible to use the tables directly to compute as many values of the power function as desired.

When $n_1 = 12$, $n_2 = 8$, and $n_1 = 15$, $n_2 = 10$, the power functions both have minima at approximately $a = 1.1$, indicating that the bias is certainly not serious. When $n_1 = n_2$, the power function has the same value at a and $1/a$, in the other cases the values shift slightly. Note that when $n_1 = n_2 = 20$ the test is fairly delicate. For example, $f(3) = .65$, that is, if $\sigma_2 = \sqrt{3}\,\sigma_1$, the probability of rejecting H_1 is .65. In Figure 1, the power functions have been plotted against log a, because of the symmetry in the values a and $1/a$. The curves I_1, I_2, I_3, correspond to columns 1, 2, 3 respectively of Table I. Similarly, curves II_1, II_2 correspond to columns 1 and 2 of Table II.

REFERENCES

[1] J. Neyman and E. S. Pearson, "On the Problem of k Samples," *Bull. Acad. Polonaise Sci. Let. (1931)*.

[2] J. Neyman and E. S. Pearson, "Contributions to the Theory of Testing Statistical Hypotheses," *Statistical Research Memoirs Vol. I (1936)*.

[3] P. P. N. Nayer, "An Investigation into the Application of Neyman and Pearson's L_1 Test," *Statistical Research Memoirs Vol. I (1936)*.

[4] S. S. Wilks and Catherine M. Thompson, "The Sampling Distribution of the Criterion λ_{H_1} when the Hypothesis Tested is not true," *Biometrika Vol. XXIX (1937)*.

[5] J. Neyman, "Tests of Statistical Hypotheses which are unbiassed in the Limit," *Annals of Math. Statistics Vol. IX no. 2.*

PRINCETON UNIVERSITY,
PRINCETON, N. J.

[5] Neyman, [5], discusses the similar property of being "unbiassed in the limit."

COMPLETE SIMULTANEOUS FIDUCIAL DISTRIBUTIONS

By M. S. Bartlett

1. **Introduction.** In a recent paper in these Annals, Starkey [13] has made some investigation of the distribution[1] related to the Behrens-Fisher test of the difference between two means from normal populations with unequal variances. She does not, however, give any critical discussion of the validity of this proposed test in the light of criticisms that have been made of it. It may therefore be an appropriate opportunity of reviewing the theory of fiducial distributions, as I see it, up to the present stage of development,[2] and in particular, of referring to the idea of complete simultaneous fiducial distributions. In conclusion I have made some brief comment on the particular problem at issue, in the light of this general theory; and have added a note on the use of approximate tests.

2. **Fiducial Probability.** If from a sample denoted symbolically by S a statistic T is obtained whose chance distribution depends on one unknown parameter θ, the distribution of T being of the form

$$p(T \mid \theta) = f(T, \theta)\, dT,$$

and if the values of T bear a regular increasing relationship with θ, (for an assigned value of the probability integral), then for any particular value $T = T_0$, we may assert that $\theta \geq \theta_0$, where

$$\int_{-\infty}^{T_0} p(T \mid \theta_0) = 1 - \epsilon,$$

and we shall know that this assertion, in the system of inferences based on the above rule, will have an exact and known probability of being wrong, given by ϵ.

The inference is thus an uncertain one, but the extent of the uncertainty is exactly known, and as stressed by Fisher [6], who first introduced this important concept of fiducial inferences and fiducial probability, is completely independent of any *a priori* notion of what value θ is likely to be.

It might be emphasized, to avoid confusion, that the inference is a *deduction* from the standpoint of logic, and still requires, if applied in practice, the necessity of inductive assumptions concerning the applicability of the mathematical theory, but its avoidance of any appeal to *a priori* probability in regard to the value of θ gives it a completely independent status district from the classical inverse probability argument, from which it should be distinguished. The

[1] This distribution has also been studied by Sukhatme [14].

[2] See also the recent expository article by Wilks [16].

interval assigned to θ is to some extent arbitrary, and we can more generally choose θ_0 and θ_1 such that the fiducial probability of

$$\theta_1 \leq \theta \leq \theta_0$$

is equal to $1 - \epsilon$. While this fiducial probability is a probability in a formal mathematical sense, I have suggested [2] that its special meaning in regard to the inference on θ might be emphasized if we distinguished it by a special symbol. Since intervals (θ_1, θ_0) can be built up for all values of ϵ, we can represent them all by the general distributional expression

$$\int_{-\infty}^{\theta} f_p(\theta \mid T) = \int_{T}^{\infty} p(T \mid \theta),$$

which defines the fiducial probability distribution $f_p(\theta \mid T)$.

From the point of view of mathematical theory T is, so far, any statistic, but Fisher restricted the term fiducial probability for those cases where T was a sufficient statistic for θ, in order that the fiducial inference should be based on a sample statistic which could justifiably claim to contain all the information on θ available from the sample.

The general theory of interval estimation, without this restriction, has been subsequently examined by Neyman (e.g. [10]) under the name of the theory of confidence intervals. In this general theory there is no particular restriction on the number of parameters involved, for it may be possible in the coordinate space represented by parameters θ_r (for which there are statistics T_s) to define a region $R(T_s)$ for which the assertion that the vector parameter θ_r lies in the region $R(T_s)$ has a known probability $1 - \epsilon$ of being correct.

A difficulty, however, in a multi-parameter theory of fiducial distributions is that it does not in general seem possible, even when T_s is a vector statistic representing a joint set of sufficient statistics for θ_r, to define a simultaneous fiducial distribution $f_p(\theta_r \mid T_s)$ which will be consistent with one-variate distributions $f_p(\theta \mid T)$ relating to one particular parameter θ. For such consistency we must have the symbolic integration

$$\int f_p(\theta_r \mid T_s)$$

over all θ_r other than θ yielding $f_p(\theta)$ as a result. A further discussion of this difficulty is given in Sections 4 and 5, after the theory of one-variate fiducial distributions has been more completely discussed.

3. **Fiducial Distributions and Properties of Sufficiency.** If we now consider the extension of one-variate fiducial inferences to the case where other parameters exist but are unknown, we are led to examine the various types of sufficient statistic which are related to the theory of estimation of one parameter when other parameters are unspecified (Bartlett, [1]). By a distribution of fiducial type we shall mean a distribution providing at least confidence inter-

vals in the sense of Neyman. This distribution will be defined as *the* fiducial distribution for θ if the statistic used (conditional or unconditional) satisfies the necessary sufficiency properties given in the paper just referred to (sections 6, p. 132, and 7, p. 136). This definition is understood to include the possibility mentioned in section 7, where $T_1 \mid T_2(\theta)$ is a conditional statistic of the type required for any specified value of θ ($T_1 \mid T_2$ denotes T_1, given T_2). For example, the theoretical statistic $\bar{x} \mid \Sigma(x - m)^2$ in normal theory, where $\bar{x}$ is the sample mean and $\Sigma(x - m)^2$ the sum of squares of deviations from the population mean m, is of this form, and since

$$p(\bar{x} \mid \Sigma(x - m)^2) = p(t),$$

a fiducial distribution for m is obtained from the familiar Student's t-distribution.

As other developments of fiducial theory we may note (i) its application to fiducial inferences on sufficient statistics in unknown samples (this application to *normal* samples has been discussed by more than one writer, see, for example, Fisher [8]; I have moreover indicated the general theory underlying such applications [3]) (ii) the case of discontinuous or "discrete" sampling distributions, for which the theory of exact fiducial distributions breaks down.

In the latter case, it is only possible to choose an interval for θ, such that the chance of our fiducial inference being incorrect is *not greater than* ϵ (see, for example, Clopper and Pearson [5]). This "inexact theory" I have shown [3] may also be extended to inferences on sufficient statistics in unknown samples. In particular, from the general distribution

$$p(r_1, r_2 \mid r) = \frac{n_1!}{(n_1 - r_1)!\,r_1!} \frac{n_2!}{(n_2 - r_2)!\,r_2!} \frac{(n - r)!\,r!}{n!}$$

giving the number of ways of assigning r members with some attribute A in numbers r_1 and r_2 to samples S_1 and S_2, sizes n_1 and n_2, we have for $n_2 = 1$,

$$p(r_2 \mid r) = \begin{cases} \dfrac{r}{n_1 + 1} & (r_2 = 1) \\ \dfrac{n_1 + 1 - r}{n_1 + 1} & (r_2 = 0). \end{cases}$$

Thus if S_1 contains r_1 members with a certain attribute, such that

$$\frac{r_1 + 1}{n_1 + 1} \leq \epsilon,$$

we may assert that a new member from the same population will not possess the attribute. If

$$\frac{n_1 + 1 - r_1}{n_1 + 1} \leq \epsilon,$$

we assert that the new member will possess the attribute. If r_1 does not conform to either inequality, we cannot, with the limit of error imposed, commit our-

selves. The probability that our variable assertion, based on the above rule, is wrong, is then not greater than ϵ. (This type of inference may be contrasted with the Law of Succession in the theory of inverse probability. In this rather degenerate example it is not of course surprising that the nature of the inference we can make is not always very profound!)

4. **Simultaneous Fiducial Distributions.** It was pointed out in section 2 that an inference of fiducial type might be made regarding a joint interval containing unknown parameters θ_r, this interval or region being a variable function of the (continuous) statistics T_s. If a sufficient set of statistics T_r $(r = 1 \cdots k)$ exist for the parameters θ_r $(r = 1 \cdots k)$, that is, if we have

$$p(S \mid \theta_r) = p(T_r \mid \theta_r)p(S \mid T_r)$$

where T_r denotes the set $T_1 \cdots T_k$, and similarly for θ_r; and if we can write

$$p(T_r \mid \theta_r) = p(\varphi_r)$$

where the distribution of the set of theoretical functions φ_r of T_r and θ_r is independent to any further extent of θ_r, then we may write also

$$f_p(\theta_r \mid T_r) = p(\varphi_r)$$

as the simultaneous fiducial distribution of the θ_r (cf. Fisher, [8]). This notation allows implicitly for the formal transformation from one set of variates to another, the last equation meaning that $p(\varphi_r)$ provides the fiducial distribution of the θ_r, when it is regarded as a distribution in θ_r. For the equations to hold, however, the Jacobian of the transformations must not change sign anywhere in the sample space, this condition determining both the formal identity of the two sides of the equations and also the necessary one-to-one relationship between values of θ_r and T_r.

It has been shown by Segal [12] that if the sufficient set T_r exist, the functions φ_r also exist. For we may define φ_r by the equations

$$\begin{cases} \varphi_1 = \displaystyle\int_{-\infty}^{T_1} p(T_1) \\ \varphi_2 = \displaystyle\int_{-\infty}^{T_2} p(T_2 \mid T_1) \\ \vdots \end{cases}$$

so that

$$p(T_1)p(T_2 \mid T_1) \cdots = d\varphi_1 d\varphi_2 \cdots \qquad (\varphi_r, 0 - 1).$$

The above theory is also immediately applicable to quasi-sufficient statistics, it being merely necessary to consider the appropriate conditional distributions.

5. **Complete Simultaneous Fiducial Distributions.** It has been emphasized [3] that the simultaneous fiducial distribution $f_p(\theta_r \mid T_r)$ obtained from a suffi-

cient set of statistics must not be interpreted analogously to a simultaneous distribution $p(T_r \mid \theta_r)$. For example, if the set T_r represent the sufficient statistics $\bar{x}$ and s^2 for the unknown mean and variance of a normal population we have

$$p(\bar{x}, s^2 \mid m, \sigma^2) = p\left(\frac{\bar{x} - m}{\sigma}, \frac{s^2}{\sigma^2}\right)$$
$$= f_p(m, \sigma^2 \mid \bar{x}, s^2),$$

but this does not imply that a fiducial inference could be made for one unknown parameter defined by $\theta = m + \sigma$ by integration of the above fiducial distribution after formal change of variable.

We may, however, in certain cases show that consistency relations are satisfied which justify to a much further extent our calling $f_p(\theta_r \mid T_r)$ a simultaneous fiducial distribution. Unfortunately this last expression has already been appropriated for $f_p(\theta_r \mid T_r)$ in general; we shall therefore call $f_p(\theta_r \mid T_r)$ a *complete* simultaneous fiducial distribution if (taking $k = 2$ for simplicity)

$$f_p(\theta_1, \theta_2) = f_p(\theta_1 \mid \theta_2) f_p(\theta_2)$$
$$= f_p(\theta_2 \mid \theta_1) f_p(\theta_1),$$

where the fiducial distributions on the right are known to exist, and their form determined, from the theory of one-variate fiducial distributions. For example, if we consider again the normal sample, we have

$$p\left(\frac{\bar{x} - m}{\sigma}, \frac{s^2}{\sigma^2}\right) = p\left(\frac{\bar{x} - m}{\sigma}\right) p\left(\frac{s^2}{\sigma^2}\right)$$
$$= f_p(m \mid \sigma^2) f_p(\sigma^2)$$

and also

$$= p\left(\frac{\bar{x} - m}{\sqrt{\Sigma}}\right) p\left(\frac{\Sigma}{\sigma^2}\right)$$
$$= f_p(m) f_p(\sigma^2 \mid m)$$

where $\Sigma \equiv \Sigma(x - m)^2$.

These relations imply not only that a fiducial region for m and σ^2 can be determined from the observed values of $\bar{x}$ and s^2, but that in particular, the region can be chosen so that (i) it is some section of an area bounded by two lines parallel to the m axis (ii) alternatively it is some section of an area bounded by two lines parallel to the σ^2 axis. Integration for m and σ^2 respectively then implies extending these sections until the whole area bounded by these two parallel lines is included in the chosen region. This existence of a complete simultaneous fiducial distribution for the two population parameters corresponding to a normal sample is a special case of the complete fiducial distribution

which exists for the two parameters of location and scaling for a sample from any population of the form

$$p(x \mid m, \sigma) = f\left(\frac{x - m}{\sigma}\right)\frac{dx}{\sigma},$$

as I have previously pointed out ([2], p. 564).[3]

For let T_1 and T_2 be any two algebraically independent statistics giving information on the two parameters, such that

$$p(S \mid m, \sigma) = p(T_1, T_2 \mid C, m, \sigma)p(C)$$

where C represents the configuration of the sample (the idea of specifying the configuration C was first introduced by Fisher [7]). The above equation is always possible, for if x_1 is the smallest observation, x_2 the next smallest and so on, let

$$\begin{aligned} T_1 &= x_1 \\ T_2 &= x_2 - x_1 \\ T_r &= (x_r - T_1)/T_2, \qquad (r > 2). \end{aligned}$$

Then $C \equiv (T_r)$ is independent of m and σ, and the quasi-sufficient set T_1, T_2 will determine a simultaneous fiducial distribution for m and σ, (the Jacobian $J\left(\frac{\varphi_1, \varphi_2}{m, \sigma}\right)$, where $\varphi_1 = \frac{T_1 - m}{\sigma}$, $\varphi_2 = \frac{T_2}{\sigma}$, is $\frac{T_2}{\sigma^3}$, and is always positive).

As further necessary conditions for $f_p(m, \sigma)$ to be complete, we have the relations

$$\begin{aligned} p(T_1, T_2 \mid C, m, \sigma) &= p(T_1 - m \mid \sigma, T_2, C)p(T_2 \mid C, \sigma) \\ &= p\left(\frac{T_1 - m}{T_2}\,\middle|\, C\right)p\left(T_1 - m \mid C, \frac{T_1 - m}{T_2}, \sigma\right). \end{aligned}$$

The first of these relations is obvious, and since the first factor in it corresponds to the quasi-sufficient statistic T_1 for m when the configuration $C' \equiv (C, T_2)$ is given (σ known), we have

$$f_p(m, \sigma) = f_p(m \mid \sigma)f_p(\sigma).$$

For the second relation we note that the set

$$\frac{T_1 - m}{T_2} \quad \text{and} \quad T_1 - m$$

are algebraically equivalent to the set T_1 and T_2. Moreover, $\frac{T_1 - m}{T_2}$ is inde-

[3] Cf. also Pitman [11], who does not, however, consider the point with which I am concerned in this paper.

pendent of σ; and if m is known, $(T_1 - m) \mid C''$, where $C'' \equiv \left(C, \frac{T_1 - m}{T_2}\right)$, is a quasi-sufficient statistic for σ. Hence

$$f_p(m, \sigma) = f_p(m) f_p(\sigma \mid m),$$

which is the relation required.

The theory of complete simultaneous fiducial distributions may be applied to sufficient statistics in unknown samples. In particular, a complete distribution may be shown to exist for the statistics $\bar{x}_2$ and s_2^2 in an unknown normal sample S_2, or for the statistics $\bar{x}$ and s^2 for the joint sample S of which the known sample S_1 is also a part ([3]; cf. Fisher, [8]).

6. **The Behrens-Fisher Test between two means.** Fisher [8] showed that by integrating out the simultaneous fiducial distribution $f_p(m, \sigma^2)$ obtained from a normal sample, we obtained either $f_p(m)$ or $f_p(\sigma^2)$. He then suggested that such integration was possible for any simultaneous fiducial distribution; and hence obtained a distribution apparently appropriate for testing the difference between two means from normal populations whose variances were unequal. Since I have shown that this integration can be justified for $f_p(m, \sigma^2)$ owing to the *complete* simultaneous nature of this distribution, it is clear that integration in any other problem is so far justified merely by analogy, and no statement as to its meaning in general has been given by Fisher.

To show more explicitly the extent to which the proposed solution is open to criticism, I examined in particular [2] the case where each estimated variance had only one degree of freedom. The Behrens-Fisher solution implies a fiducial distribution

$$f_p(\delta) = \frac{(s_1 + s_2)\, d\psi}{\pi\{(s_1 + s_2)^2 + \psi^2\}}$$

where δ is the difference in population means $m_1 - m_2$, $\psi = (m_1 - x_1) - (m_2 - x_2)$, where x_1 and x_2 are sample means with estimated variances s_1^2 and s_2^2 each based on only one degree of freedom. By direct argument, I derived a distribution of fiducial type

$$f_p(\delta) = \frac{|\, s_1 \pm s_2 \,|\, d\psi}{\pi\{(s_1 \pm s_2)^2 + \psi^2\}}$$

where the sign + or − is to be decided at random. It is irrelevant to my argument whether we are justified in calling this distribution *the* fiducial distribution of δ; it is also irrelevant what distribution would ensue if the + and − signs were considered separately. It is sufficient to note that the distribution certainly provides us with an exact inference of fiducial type, as Fisher himself confirmed ([9], p. 375); and this inference clashes with the apparent inference to be drawn from the Behrens-Fisher solution. In general it is of course true that different distributions might validly lead to different inferences of fiducial

type, but here the distributions are sufficiently similar mathematically for it to be possible to assert that they cannot both be correct. The direct distribution of $\psi/(s_1 + s_2)$ is in fact known to be dependent on the unknown ratio ϕ of the population variances (Fisher, [9], p. 374). While Fisher suggests that this in no way invalidates his fiducial argument, in my view if an inference is to be independent of an unknown parameter, it should in particular be independent of it if we imagine that we are being supplied with pairs of samples, for all of which the ratio ϕ has the same value.

7. **Approximate Tests.** I have shown ([2], p. 565) that while $f_p(\delta)$ in general does not appear to exist, we have

$$f_p(\delta, \phi) = f_p(\delta \mid \phi) f_p(\phi)$$

where

$$f_p(\delta \mid \phi) = C\left\{1 + \frac{\psi^2\phi}{(1+\phi)(n_1 s_1^2 + n_2 s_2^2 \phi)}\right\}^{-\frac{1}{2}(n_1+n_2+1)} \cdot d\left\{\frac{\psi}{\sqrt{(n_1 s_1^2 + n_2 s_2^2\phi)}} \sqrt{\frac{\phi}{1+\phi}}\right\}$$

where n_1 and n_2 are the degrees of freedom of s_1 and s_2, and C is a constant. For $n_1 = n_2$, the fiducial limits for δ (if ϕ were known) were shown to be insensitive to changes in ϕ, as has also been shown by Welch [15] in more detail. For $n_1 \neq n_2$ this is no longer the case. If we tried to get an approximate solution we might consider inserting $\theta = s_1^2/s_2^2$ for ϕ in the above distribution; this would be equivalent to considering the (direct) distribution of

$$T = \frac{x_1 - x_2}{\sqrt{(s_1^2 + s_2^2)}}$$

as a t-distribution with $n_1 + n_2$ degrees of freedom. This is therefore a first approximation to the true distribution of T, which has been obtained by Welch [15] to a further approximation involving ϕ.

Sometimes it is sufficient in practice if we can assign limits to the true significance level of T in any problem, as was illustrated in my own paper ([2], p. 566). A formal proof of the inequality used there is as follows.

The actual distribution of T for $n_1 = n_2 = n$, say, depends on the integral

$$I(\phi) = \int_0^\infty \frac{\lambda}{(1 + \lambda^2 T^2/n)^{n+\frac{1}{2}}} \frac{(\theta/\varphi)^{\frac{1}{2}n-1} d(\theta/\varphi)}{(1+\theta/\varphi)^n}$$

where

$$\lambda^2 = \frac{2\varphi(1+\theta)}{(1+\varphi)(\theta+\varphi)}$$

and hence the significance level of T on the integral

$$J(\phi) = \int_0^{|T|} I(\varphi)\, dT.$$

If we write

$$u = \begin{cases} \theta/\varphi, & (0 \leq \theta/\varphi \leq 1) \\ \varphi/\theta, & (1 \leq \theta/\varphi < \infty) \end{cases}$$

we obtain

$$J(\phi) = \int_0^{|T|} \int_0^1 \left\{ \frac{\lambda_1}{(1 + \lambda_1^2 T^2/n)^{n+\frac{1}{2}}} + \frac{\lambda_2}{(1 + \lambda_2^2 T^2/n)^{n+\frac{1}{2}}} \right\} \frac{u^{\frac{1}{2}n-1}}{(1+u)^n} dT,$$

that is,

$$\lambda_1^2 = \frac{2(1+u\varphi)}{(1+u)(1+\varphi)}, \qquad \lambda_2^2 = \frac{2(u+\varphi)}{(1+u)(1+\varphi)};$$

where

$$J(\varphi) = \int_0^1 \{F(t_1) + F(t_2)\} u^{\frac{1}{2}n-1}(1+u)^{-n} du$$

where $t_1 = \lambda_1 |T|$, $t_2 = \lambda_2 |T|$, and $F(t)$ is proportional to the probability integral of a t with n degrees of freedom. Since

$$\frac{\partial(F(t_1) + F(t_2))}{\partial \varphi} \propto \left\{ \frac{1}{\lambda_2(1 + \lambda_2^2 T^2/n)^{n+\frac{1}{2}}} - \frac{1}{\lambda_1(1 + \lambda_1^2 T^2/n)^{n+\frac{1}{2}}} \right\} \frac{(1-u)|T|}{(1+u)(1+\varphi)^2}$$

this differential coefficient, from the relations

$$(1-u)(1-\varphi) \geq 0,$$

$$1 + u\varphi \geq u + \varphi,$$

$$\lambda_1 \geq \lambda_2,$$

is never negative for all u and $\varphi(\varphi \leq 1)$. Hence $J(\varphi)$ is a steadily increasing function in the range (0, 1) for all values of T; or the significance level of T lies between its values for $\varphi = 0$ and $\varphi = 1$, as previously stated.

More generally, for $n_1 \neq n_2$, the effective number of degrees of freedom for T would be expected to lie between n_1 $(n_1 \leq n_2)$ and $n_1 + n_2$ (cf. Welch, [15], p. 360), though I have not succeeded in establishing this rigorously by a modification of the above proof.

REFERENCES

[1] M. S. Bartlett, *Proc. Roy. Soc.* Vol. A154 (1936), pp. 124–137.
[2] M. S. Bartlett, *Proc. Camb. Phil. Soc.* Vol. 32 (1936), pp. 560–566.
[3] M. S. Bartlett, *Proc. Roy. Soc.* Vol. A160 (1937), pp. 268–282.
[4] W. V. Behrens, *Landw. Jb.* Vol. 68 (1929), pp. 807–837.
[5] C. J. Clopper and E. S. Pearson, *Biometrika*, Vol. 26 (1934), pp. 404–413.
[6] R. A. Fisher, *Proc. Camb. Phil. Soc.* Vol. 26 (1930), pp. 528–535.
[7] R. A. Fisher, *Proc. Roy. Soc.* Vol. A144 (1934), pp. 285–307.
[8] R. A. Fisher, *Ann. Eugen.* Vol. 6 (1935), pp. 391–398.
[9] R. A. Fisher, *Ann. Eugen.* Vol. 7 (1937), pp. 370–375.

[10] J. Neyman, *Phil. Trans.* Vol. A236 (1937), pp. 333–380.
[11] E. J. Pitman, *Biometrika* Vol. 30 (1939), pp. 391–421.
[12] I. E. Segal, *Proc. Camb. Phil. Soc.* Vol. 34 (1938), pp. 41–47.
[13] D. M. Starkey, *Ann. Math. Stat.* Vol. 9 (1938), pp. 201–213.
[14] P. V. Sukhatme, *Sankhya* Vol. 4 (1938), pp. 39–48.
[15] B. L. Welch, *Biometrika* Vol. 29 (1938), pp. 350–361.
[16] S. S. Wilks, *Ann. Math. Stat.* Vol. 9 (1938), pp. 272–280.

CAMBRIDGE UNIVERSITY,
ENGLAND

ON TESTS OF SIGNIFICANCE IN TIME SERIES

By G. Tintner

The purpose of this note is to give some tests of significance for problems connected with time series. H. Wold [1], in a recent book gives an excellent theoretical treatment of this subject (without treating, however, the important problem of the trend), but he does not give any tests of significance, [2]. These have proved extremely important in other fields, especially in biological applications. The method used in what follows may be found useful also for other problems in time series.

1. **A Test of Significance for the Variances of Differences.** The Variate Difference Method, [3, 4, 5], starts from the assumption that the time series w_i $(i = 1, 2 \cdots N)$ consists of two additive parts: A "smooth" part m_i, the mathematical expectation of w_i, and a random part x_i, which we will assume to be normally and independently distributed with mean 0 and variance σ^2. Hence we have

$$w_i = m_i + x_i, \qquad i = 1, \cdots, N \tag{1}$$

if we have N items in our series.
We form the finite differences and get for the difference of order k

$$\Delta^k w_i = \Delta^k m_i + \Delta^k x_i. \tag{2}$$

But the smooth component or mathematical expectation can be eliminated to any desired degree by successive differencing. This would not be true of a "zig-zag" component or a periodic function with small period [6]. It will be remembered, that for instance the differences of order k of a polynomial of order k are constant and that differences of order $(k + 1)$ and higher are zero.

O. Anderson and others who worked in this field have tested the order of the difference, say k_0, beginning from where the component m_i is sufficiently eliminated, in the following way. They define the variance σ_k^2 of the k-th difference by

$$\sigma_k^2 = \sum_{i=1}^{N-k} (\Delta^k w_i)^2 / ((N - k)_{2k} C_k). \tag{3}$$

We note that all variances of the differences beginning from k_0 must be equal to each other, because they will contain only the component x_i, if the other component m_i has been eliminated through taking k_0 differences. O. Anderson and R. Zaykoff, [3, 7], give formulae for the standard errors of the difference

between the variances of the differences k and $k + 1$. These formulae are valid only for large samples and suppose a knowledge of the true variance σ^2.

We propose a new method for testing the equality of the variances of two successive differences in order to find the order of the difference k_0 beginning from which we have

$$\sigma^2_{k_0} = \sigma^2_{k_0+1} = \sigma^2_{k_0+2} = \cdots \tag{4}$$

This method is one of selection and it consists in selecting the items to be included in the variance of the k-th difference, σ^2_k, in such a way that they become independent of the items to be included in the calculation of the variance of the difference of order $k + 1$, σ^2_{k+1}. Then the ordinary test of significance, i.e. the one involved in the analysis of variance as developed by R. A. Fisher [8], becomes applicable.

Let us consider an example. Suppose we want to compare the variance of the first differences and of the second differences, in order to test the hypothesis that the component m_i has already been eliminated in the first difference. But the process of forming finite differences has introduced correlations, even if the original random elements x_i are independently distributed. Each item in the series of the first differences will be correlated with the next and the preceding item. Each item in the series of second differences will be correlated with the two preceding and the two following items of the same series. But each item of the series of the first differences will also be correlated with the two preceding, the corresponding and the following item of the series of second differences.

We can make a very simple valid comparison in spite of these correlations if we sacrifice some of the available information. We can for instance calculate $\sigma_1^{2'}$ by including only items number 1, 6, 11, 16 etc. of the series of first differences. And we calculate $\sigma_2^{2'}$ by including only items number 3, 8, 13, 18 etc. of the series of second differences. The two quantities $\sigma_1^{2'}$ and $\sigma_2^{2'}$ are independent and hence can be compared by using either Fisher's z test, [8], or Snedecor's F table [9]. The variances are

$$\sigma_1^{2'} = \sum{}' (\Delta^1 w_i)^2 \Big/ \left(\frac{(N-1)}{5}\, {}_2C_1\right) \quad \text{and} \tag{5}$$

$$\sigma_2^{2'} = \sum{}'' (\Delta^2 w_i)^2 \Big/ \left(\frac{(N-2)}{5}\, {}_4C_2\right) \tag{6}$$

where $\sum'$ and $\sum''$ denote summation over the selected items. Other selections which are possible are: Items number 2, 7, 12 etc. of the series of first differences and items number 4, 9, 14 etc. of the series of second differences. Or items number 3, 8, 13 of the series of the first differences and items number 5, 10, 15 etc. of the series of second differences. Or items number 4, 9, 14 etc. of the series of first differences and items number 6, 11, 16 etc. of the series of second differences. Finally, items number 5, 10, 15 etc. of the series of first

differences and items number 7, 12, 17 etc. of the series of second differences. These 5 selections are of course not independent of each other. The comparison can always be made by calculating the variances according to formulae (5) and (6) and using either Fisher's z table, [8], or Snedecor's F table, [9], for $(N-1)/5$ and $(N-2)/5$ degrees of freedom. If N is large enough, these two numbers will be near enough together in order to use the property of the z distribution to become normal for equal degrees of freedom, [8]. Then we can assume that $z = (\log {}_e\sigma_1^{2'} - \log {}_e\sigma_2^{2'})/2$ is normally distributed with mean zero and standard error $\sqrt{5/(N-2)}$.

Should the test turn out positive, i.e. if the difference between the variances is greater than permitted from the point of view of certain significance levels, then we have to compare the variance of the second and the third differences, by selecting items in a similar manner and so on.

The general procedure is as follows: If we want to compare the variance of the difference number k and the difference number $k+1$, we find that we can only use a part of our available series, because we must make a selection in order to get two independent estimates. We can make $2k+3$ different selections, which are not independent but each give two unbiased, independent estimates of the variances of the differences k and $k+1$. The selections consist in taking items number j, $j+(2k+3)$, $j+2(2k+3)$, $j+3(2k+3)$ etc. of the series of k-th differences and items number $j+k+1$, $j+k+1+(2k+3)$, $j+k+1+2(2k+3)$, $j+k+1+3(2k+3)$ etc. of the series of $(k+1)$ difference. j is here equal to 1, 2, 3 $\cdots$ $2k+3$, giving $2k+3$ possible selections for the comparison.

The variances of the difference number k and $k+1$ are calculated according to the formulae

$$\sigma_k^{2'} = \sum{}'(\Delta^k w_i)^2 \Big/ \left(\frac{(N-k)}{2k+3}\, {}_{2k}C_k\right) \tag{7}$$

$$\sigma_{k+1}^{2'} = \sum{}''(\Delta^{k+1} w_i)^2 \Big/ \left(\frac{(N-k-1)}{2k+3}\, {}_{2k+2}C_{k+1}\right). \tag{8}$$

The summations are again taken over the selected items and we can make an ordinary analysis of variance with Fisher's z table or Snedecor's F table entering it for $(N-k)/(2k+3)$ and $(N-k-1)/(2k+3)$ degrees of freedom. If N is appreciably large, we can assume the number of degrees of freedom as equal and $z = (\log {}_e\sigma_k^{2'} - \log {}_e\sigma_{k+1}^{2'})/2$ is normally distributed about zero with a standard error of $\sqrt{(2k+3)/(N-k-1)}$.

2. The Distribution of the Serial Co-variance. A similar method yields the distribution of the serial covariance, i.e., the product of a random series with itself if lagged by a lag L. We assume that x_i, $i = 1, \cdots, N$, is a series of N terms which are normally and independently distributed with mean zero and variance one. We form the serial covariance w by lagging it by L terms and

make a selection. We only include the products number 1, $1 + (L + 1)$, $1 + 2(L + 1)$, $1 + 3(L + 1)$ etc. The following formulae are exact only if N is a multiple of $L + 1$, otherwise they have to be regarded as approximations. The serial covariance w is

$$w = (x_1x_{L+1} + x_{L+2}x_{2L+2} + x_{2L+3}x_{3L+3} + \cdots) \Big/ \left(\frac{N}{L+1}\right) \tag{9}$$

We shall use the method of characteristic functions, [10], in order to establish the distribution of w. The characteristic function is in our case

$$E(e^{iwy}) = g(y) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} e^{iwy} f\, dx_1 \cdots dx_N \tag{10}$$

where f is the distribution function of the x_i, i.e., a distribution of N normal and independent variates with zero means and unit variances.

An orthogonal transformation of the quadratic form in the exponent yields a determinant, which consists of $\dfrac{N}{L+1}$ steps each of the form

$$\begin{vmatrix} 1 & 0 & \cdots & -i(L+1)y/N \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ -i(L+1)y/N & 0 & & 1 \end{vmatrix} = 1 + \frac{(L+1)^2y^2}{N^2}. \tag{11}$$

The characteristic function is therefore given by

$$g(y) = \left[1 + \frac{(L+1)^2y^2}{N^2}\right]^{-\frac{1}{2}N(L+1)} \tag{12}$$

and the distribution of w, say $D(w)$, is given, [11], by

$$D(w) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iwy} g(y)\, dy$$

$$= \frac{2^{\frac{1}{2}\left(2-\frac{N}{L+1}\right)}\left(\frac{N}{L+1}\right)^{\frac{1}{2}\left(1+\frac{N}{L+1}\right)}}{\sqrt{2\pi}\,\Gamma\left(\frac{N}{2L+2}\right)} w^{\frac{1}{2}\left(\frac{N}{L+1}-1\right)} K_{\frac{1}{2}\left(\frac{N}{L+1}-1\right)}\left(\frac{Nw}{L+1}\right) \tag{13}$$

where K is a Bessel function of the second kind for a purely imaginary argument, [12].

We can also get from (12) an asymptotic formula for large N. In this case w is distributed normally about zero with a variance of $(L + 1)/N$.

REFERENCES

[1] H. Wold, *Analysis of Stationary Time Series*, Upsala, 1938.

[2] M. S. Bartlett, "Some Aspects of the Time Correlation Problem in Regard to Tests of Significance," *Journal of the Royal Statistical Society*, vol. 98, 1935, pp. 536 ff.

[3] O. Anderson, "On the Logic of Decomposition of Statistical Series into Separate Components," *Journal of the Royal Statistical Society*, vol. 90, 1927, pp. 548 ff.
[4] O. Anderson, *Die Korrelationrechnung in der Konjunkturforschung*, Bonn, 1929.
[5] G. Tintner, *Prices in the Trade Cycle*, Vienna, 1935.
[6] A. Wald, *Berechnung und Ausschaltung von Saisonschwankungen*, Vienna, 1936, pp. 37 ff.
[7] R. Zaykoff, "Ueber die Ausschaltung der zufaellingen Komponente nach der Variate Difference Methods," *Publications of the Statistical Institute for Economic Research*, State University of Sofia, 1937, no. 1, pp. 75 ff.
[8] R. A. Fisher, *Statistical Methods for Research Workers*, 4th ed., London, 1932, pp. 202 ff.
[9] G. W. Snedecor, *Statistical Methods*, Ames, Iowa, 1938, pp. 184 ff.
[10] S. S. Wilks, *Statistical Inference*, Princeton, 1937, pp. 34 ff.
[11] G. N. Watson, *The Theory of Bessel Functions*, Cambridge, 1922, p. 172.
[12] M. S. Bartlett and J. Wishart, "The Distribution of Second Order Moment Statistics in a Normal System," *Proceedings Cambridge Philosophical Society*, vol. 28, 1931, pp. 455 ff.

IOWA STATE COLLEGE,
AMES, IOWA.

ON AN INTEGRAL EQUATION IN POPULATION ANALYSIS

By Alfred J. Lotka

I

A fundamental equation in population analysis rests on the following considerations: Of the persons born a years ago a certain fraction $p(a)$, ascertainable for example by means of a life table, survives to age a, and forms the a-year-old contingent of the existing population. A similar remark applies to every age of life. If, therefore, we denote by $N(t)$ the number of the population at time t, and by $B(t)$ the annual rate of births at the same time, and if we are dealing with a *closed* population, that is, one exempt from immigration and emigration, then, evidently,

$$N(t) = \int_0^\infty B(t-a)p(a)\,da. \tag{1}$$

In general $p(a)$ may be a function of t also, but we shall here consider primarily the case where $p(a)$ does not contain t explicitly.

The function $p(a)$ being known (from a life table), if $B(t)$ is given as a function of t, then $N(t)$ follows by direct integration of the right hand member of (1).

If, on the contrary, $N(t)$ is given, and $B(t)$ is to be determined, a special problem arises. On a former occasion[1] I have given a solution for cases in which the function $N(t)$ is given or can be expanded in the form of a series proceeding in ascending powers of e^{rt}, where r is constant; and, more particularly, for the case in which $N(t)$ is the logistic function

$$N(t) = \frac{N_\infty}{1+e^{-rt}} = N_\infty(e^{rt} - e^{2rt} + e^{3rt} - \cdots). \tag{2}$$

Although $N(t)$ is expanded in an exponential series in the process of obtaining the solution by this method, in the final result these terms are reunited, and only the original function $N(t)$ as such, together with its derivatives, appears. This suggests that it should be possible to obtain the result by a more direct route, retaining the function in its original form throughout the process. This is indeed the case, as will now be shown, by a method which at the same time frees us from the assumption that $N(t)$ can be represented by an exponential series in powers of e^{rt}.

This is accomplished as follows:

---

[1] A. J. Lotka, *Proc. Natl. Acad. Sci.*, 1929, vol. 15, p. 793; *Human Biology*, 1931, vol. 3, p. 459.

Let us put

$$N(t) = \varphi_0(t) \tag{3}$$

and assume that $B(t)$ can be expressed as a series in $\varphi_0(t)$ and its derivatives, thus

$$B(t) = c_0\varphi_0(t) - c_1\varphi_1(t) + \frac{c_2}{2!}\varphi_2(t) - \frac{c_3}{3!}\varphi_3(t) + \cdots \tag{4}$$

$$\begin{aligned} B(t-a) = {} & c_0\left\{\varphi_0(t) - a\varphi_1(t) + \frac{a^2}{2!}\varphi_2(t) - \frac{a^3}{3!}\varphi_3(t) + \cdots\right\} \\ & - c_1\left\{\qquad \varphi_1(t) - a\ \varphi_2(t) + \frac{a^2}{2!}\varphi_3(t) - \cdots\right\} \\ & + \frac{c_2}{2!}\left\{\qquad\qquad \varphi_2(t) - a\ \varphi_3(t) + \cdots\right\} \\ & - \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{aligned} \tag{5}$$

where $\varphi_n(t)$ denotes the nth derivative of $\varphi_0(t)$.

Introducing (5) in (1), and carrying out the integration, we obtain

$$\begin{aligned} \varphi_0(t) = {} & c_0m_0\varphi_0(t) - \{c_1m_0 + c_0m_1\}\varphi_1(t) + \frac{1}{2!}\{c_2m_0 + 2c_1m_1 + c_0m_2\}\varphi_2(t) \\ & - \frac{1}{3!}\{c_3m_0 + 3c_2m_1 + 3c_1m_2 + c_0m_3\}\varphi_3(t) + \cdots \end{aligned} \tag{6}$$

where m_n denotes the nth moment of the function $p(a)$ about the origin of a, that is,

$$m_n = \int_0^\infty a^n p(a)\,da. \tag{7}$$

Equation (6) is satisfied by putting

$$\begin{cases} 1 = c_0m_0 \\ 0 = c_0m_1 + c_1m_0 \\ 0 = c_0m_2 + 2c_1m_1 + c_2m_0 \\ 0 = c_0m_3 + 3c_1m_2 + 3c_2m_1 + c_3m_0 \end{cases} \tag{8}$$

the numerical coefficients being those of the corresponding binomial expansion.

Now consider

$$\begin{aligned} \beta(r) = \frac{1}{\displaystyle\int_0^\infty e^{-ra}p(a)\,da} &= \frac{1}{m_0 - m_1r + \dfrac{m_2}{2!}r^2 - \cdots} \\ &= C_0 - C_1r + \frac{C_2}{2!}r^2 - \cdots \end{aligned} \tag{9}$$

This gives

$$(10)\qquad \begin{cases} 1 = C_0 m_0 \\ 0 = C_0 m_1 + C_1 m_0 \\ 0 = C_0 m_2 + 2C_1 m_1 + C_2 m_0 \\ 0 = C_0 m_3 + 3C_1 m_2 + 3C_2 m_1 + C_3 m_0 \end{cases}$$

from which it is seen that the coefficients c in equation (6) are identical with the coefficients C in equation (9), that is, they are the coefficients of successive powers of r in the expansion of

$$(11)\qquad \beta(r) = \frac{1}{\int_0^\infty e^{-ra} p(a)\, da}$$

as a power series in r.

These coefficients can also be conveniently expressed in terms of the Thiele seminvariants λ of the function $p(a)$, which are defined by

$$(12)\qquad \begin{aligned} \int_0^\infty e^{-ra} p(a)\, da &= m_0 e^{-\lambda_1 r + \lambda_2 \frac{r^2}{2!} - \lambda_3 \frac{r^3}{3!} + \cdots} \\ &= m_0 - m_1 r + \frac{m_2}{2!} r^2 - \frac{m_3}{3!} r^3 + \cdots \end{aligned}$$

Differentiating the right hand member of (12) we have

$$(13)\qquad \begin{aligned} & m_0 \left(\lambda_1 - \lambda_2 r + \lambda_3 \frac{r^2}{2!} - \cdots\right) e^{-\lambda_1 r + \lambda_2 \frac{r^2}{2!} - \lambda_3 \frac{r^3}{3!} + \cdots} \\ &= \left(\lambda_1 - \lambda_2 r + \lambda_3 \frac{r^2}{2!} - \cdots\right)\left(m_0 - m_1 r + m_2 \frac{r^2}{2!} - \cdots\right) \\ &= m_1 - m_2 r + m_3 \frac{r^2}{2!} - \cdots \end{aligned}$$

Hence

$$(14)\qquad \begin{cases} m_0 = m_0 \\ m_1 = \lambda_1 m_0 \\ m_2 = \lambda_1 m_1 + \lambda_2 m_0 \\ m_3 = \lambda_1 m_2 + 2\lambda_2 m_1 + \lambda_3 m_0 \\ m_4 = \lambda_1 m_3 + 3\lambda_2 m_2 + 3\lambda_3 m_1 + \lambda_4 m_0 \end{cases}$$

again with binomial coefficients.

The λ's being thus defined, we now have

$$\frac{1}{\int_0^\infty e^{-ra}p(a)\,da} = \frac{1}{m_0}e^{\lambda_1 r - \lambda_2\frac{r^2}{2!} + \lambda_3\frac{r^3}{3!} - \cdots} \tag{15}$$

$$= c_0 - c_1 r + c_2\frac{r^2}{2!} - \cdots \tag{16}$$

from which it follows, as in the case of equations (12), (13), that

$$\begin{cases} c_0 = \dfrac{1}{m_0} \\ c_1 = -\lambda_1 c_0 \\ c_2 = -\lambda_1 c_1 - \lambda_2 c_0 \\ c_3 = -\lambda_1 c_2 - 2\lambda_2 c_1 - \lambda_3 c_0 \\ c_4 = -\lambda_1 c_3 - 3\lambda_2 c_2 - 3\lambda_3 c_1 - \lambda_4 c_0 \\ c_5 = -\lambda_1 c_4 - 4\lambda_2 c_3 - 6\lambda_3 c_2 - 4\lambda_4 c_1 - \lambda_5 c_0 \end{cases} \tag{17}$$

once more with binomial coefficients. The coefficients c are, in fact, related to the negative seminvariants $-\lambda$ in the same way as the moments m are related to the direct seminvariants.

Considerable simplification in the coefficients c can be effected by a change in origin of t. This is most easily accomplished by reverting to equation (1) in which we write, instead of $B(t-a)$, the equivalent expression

$$B(t-a) = B\{(t-\lambda_1) - (a-\lambda_1)\} = B(\theta - \alpha). \tag{18}$$

In place of the moments m of the function $p(a)$, taken about $a = 0$, there then appear in (6) the corresponding moments taken about the mean age λ_1, and in the equation corresponding to (17), for the new coefficients c_0', c_1', c_2', $\cdots$ the seminvariants λ are now defined in terms of these new moments. According to a well-known property of the Thiele seminvariants this leaves all the λ's except λ_1 unchanged, while reducing this latter to zero.

The coefficients c' are therefore obtained by a set of equations identical with those for the coefficients c, in which, however, the substitution $\lambda_1 = 0$ eliminates all terms containing either λ_1 or c_1, thus

$$\begin{cases} c_0' = c_0 = \dfrac{1}{m_0} \\ c_1' = 0 \\ c_2' = -\lambda_2 c_0 \\ c_3' = -\lambda_3 c_0 \\ c_4' = -(\lambda_4 - 3\lambda_2^2)c_0 \\ c_5' = -(\lambda_5 - 10\lambda_2\lambda_3)c_0 \end{cases} \tag{19}$$

With this choice of constants the solution (4) of the fundamental equation (1) finally takes the form

$$(20)\quad B(t) = \frac{1}{m_0}\left\{\varphi_0(\theta) - \frac{\lambda_2}{2!}\varphi_2(\theta) + \frac{\lambda_3}{3!}\varphi_3(\theta) - \frac{(\lambda_4 - 3\lambda_2^2)}{4!}\varphi_4(\theta) + \cdots\right\}.$$

It is thus seen that if the population, as a function of the time, is represented by $\varphi(t)$, and expansion by Taylor's theorem is applicable to $\varphi(t - a)$ within the range $0 < a < \omega$ where ω is the highest age attained by any individual, i.e., the highest age for which $p(a)$ has a value other than zero, then the annual births $B(t)$ under the régime of a constant life table can be represented by a series (20) proceeding in successive derivatives of φ. The constant coefficients of the successive members of the series are known functions[2] of the Thiele semi-variants λ of the function $p(a)$, the probability at birth of attaining age a.

II. ALTERNATIVE SOLUTION

1. In the special case of a population growing at a constant rate r under the régime of a constant life table, the constant birth rate per head is given by

$$(21)\quad \beta(r) = \frac{1}{\int_0^\infty e^{-ra}p(a)\,da}.$$

This suggests that when r is variable we may still have as a first approximation

$$(22)\quad \beta(t) = \frac{1}{\int_0^\infty e^{-r_t a}p(a)\,da} = \beta(r_t)$$

and that this expression may form the first term in a series expansion of some kind. Evidence of this has, indeed, been shown[3] in the case of a population growing according to the logistic curve, but the formal justification of the supposition was not fully established, nor was the law of the series expansion determined. We now proceed to establish the series for the general case, using as a starting point the result obtained in Part I.

We revert, then, to equation (4), and, dividing by $N(t)$, we have

$$(23)\quad \frac{B(t)}{N(t)} = b(t) = c_0 - c_1\frac{\varphi_1(t)}{\varphi_0(t)} + \frac{c_2\varphi_2(t)}{2!\,\varphi_0(t)} - \frac{c_3\varphi_3(t)}{3!\,\varphi_0(t)} + \cdots$$

[2] An obvious extension of this result is that this representation of $B(t)$ may still hold approximately when the life table is variable, and the seminvariants λ are accordingly functions of t. We may expect this approximation to be serviceable when $p(a, t)$ changes but slowly with t, a condition that will usually be satisfied in practice. See A. J. Lotka, *Human Biology*, 1931, vol. 3, p. 481.

[3] A. J. Lotka, *Proceedings Natl. Acad. Sci.*, 1929, Vol. 15, p. 796.

But $\frac{\varphi_1(t)}{\varphi_0(t)}$ is the rate of increase per head at time t, which we have denoted by r_t, that is

$$\varphi_1(t) = r_t\varphi_0(t). \tag{24}$$

To systematize notation, let us write r_1 instead of r_t, and denote by $r_2, r_3 \cdots$ successive derivatives of r_1 with respect to t. With this notation the following scheme, homogeneous as regards the weight of the terms in the right hand member, results

$$\begin{cases} \varphi_0 = \varphi_0 \\ \varphi_1 = r_1\varphi_0 \\ \varphi_2 = r_1\varphi_1 + r_2\varphi_0 \\ \varphi_3 = r_1\varphi_2 + 2r_2\varphi_1 + r_3\varphi_0 \\ \varphi_4 = r_1\varphi_3 + 3r_2\varphi_2 + 3r_3\varphi_1 + r_4\varphi_0 \end{cases} \tag{25}$$

again with binomial coefficients.

Eliminating derivatives of φ from the right hand members of the set of equations (25), we find

$$\begin{cases} \frac{\varphi_1}{\varphi_0} = r_1 \\ \frac{\varphi_2}{\varphi_0} = r_1^2 + r_2 \\ \frac{\varphi_3}{\varphi_0} = r_1^3 + 3r_1r_2 + r_3 \\ \frac{\varphi_4}{\varphi_0} = r_1^4 + 6r_1^2r_2 + 4r_1r_3 + 3r_2^2 + r_4 \end{cases} \tag{26}$$

Introducing the expression (26) for $\frac{\varphi_n}{\varphi_0}$ in (23), and rearranging terms, we find

$$\begin{aligned} b(t) = {} & c_0 - c_1r_1 + \frac{c_2r_1^2}{2!} - \frac{c_3r_1^3}{3!} + \cdots \\ & + \frac{r_2}{2!}\left(c_2 - c_3r_1 + \frac{c_4r_1^2}{2!} - \cdots\right) \\ & - \frac{r_3}{3!}\left(c_3 - c_4r_1 + \frac{c_5r_1^2}{2!} - \cdots\right) \\ & + \frac{(r_4 + 3r_2^2)}{4!}\left(c_4 - c_5r_1 + \frac{c_6r_1^2}{2!} - \cdots\right) \\ & - \frac{(r_5 + 10r_2r_3)}{5!}\left(c_5 - c_6r_1 + \frac{c_7r_1^2}{2!} - \cdots\right). \end{aligned} \tag{27}$$

It will be seen that the factors r_2, r_3, $(r_4 + 3r_2^2)$, $(r_5 + 10r_2r_3)$, etc., by which, in successive terms, the power series in r_1 are multiplied, are obtained by the formal substitution $r_1 = 0$ in the corresponding expressions (26) for $\frac{\varphi_n}{\varphi_0}$, as for example,

$$(28)\quad \begin{cases} \dfrac{\varphi_0}{\varphi_0} = 1 & \dfrac{\varphi_3}{\varphi_0} = r_1^3 + 3r_1r_2 + r_3 \\ \dfrac{\varphi_1}{\varphi_0} = r_1 & \left[\dfrac{\varphi_3}{\varphi_0}\right]_0 = r_3 \\ \left[\dfrac{\varphi_1}{\varphi_0}\right]_0 = 0 & \dfrac{\varphi_4}{\varphi_0} = r_1^4 + 6r_1^2r_2 + 4r_1r_3 + 3r_2^2 + r_4 \\ \dfrac{\varphi_2}{\varphi_0} = r_1^2 + r_2 & \left[\dfrac{\varphi_4}{\varphi_0}\right]_0 = 3r_2^2 + r_4 \\ \left[\dfrac{\varphi_2}{\varphi_0}\right]_0 = r_2 & \end{cases}$$

With this interpretation of the symbol $\left[\frac{\varphi_n}{\varphi_0}\right]_0$, we may therefore write

$$(29)\qquad b(t) = \sum_{n=0}^{n=\infty} \frac{1}{n!}\left[\frac{\varphi_n}{\varphi_0}\right]_0 \frac{\partial^n \beta(r_1)}{\partial r_1^n}$$

with the understanding that

$$\frac{\partial^0 \beta(r_1)}{\partial r_1^0} = \beta(r_1) = \beta(r_t).$$

Furthermore, since $\frac{\varphi_0}{\varphi_0} = 1$ and $\left[\frac{\varphi_1}{\varphi_0}\right]_0 = 0$, equation (29) can also be written in the form

$$(30)\qquad b(t) = \beta(r_t) + \sum_{n=2}^{n=\infty} \frac{1}{n!}\left[\frac{\varphi_n}{\varphi_0}\right]_0 \frac{\partial^n \beta(r_1)}{\partial r_1^n}$$

which establishes the desired result, namely, that $b(t)$ is expressed in terms of a fully defined series, in which the first term is

$$(22)\qquad \beta(r_t) = \frac{1}{\int_0^\infty e^{-r_t a} p(a)\, da}.$$

It will be observed that equation (23) can be written

$$b(t) = \sum \frac{1}{n!}\left[\frac{\partial^n \beta(r)}{\partial r^n}\right]_0 \frac{\varphi_n}{\varphi_0}$$

so that, in view of (29), we have

$$\sum \frac{1}{n!}\left[\frac{\partial^n \beta(r)}{\partial r^n}\right]_0 \frac{\varphi_n}{\varphi_0} = \sum \frac{1}{n!}\frac{\partial^n \beta(r)}{\partial r^n}\left[\frac{\varphi_n}{\varphi_0}\right]_0$$

a somewhat remarkable relation.

Analytically, our problem must thus be considered solved, but for purposes of computation, as well as on account of a certain analytical interest of their own, it is desirable to examine certain properties of the various characteristics that appear in the treatment of the problem.

2. **Successive partial derivatives of $\beta(r)$.** In the application of the formulae (29) or (30), it is necessary to obtain successive partial derivatives of $\beta(r)$ with respect to r. The values of the derivatives $\frac{\partial^n \beta}{\partial r^n}$ can be computed directly from (9), but more exact values are obtained by taking advantage of certain special properties of these derivatives. With this in view, it is desirable, first of all, to consider certain properties of the moments M_n and the seminvariants Λ_n of the function

$$f(r) = e^{-ra}p(a). \tag{31}$$

We note that

$$M_n = \int_0^\infty a^n e^{-ra} p(a)\, da \tag{32}$$

$$\frac{\partial M_n}{\partial r} = -\int_0^\infty a^{n+1} e^{-ra} p(a)\, da$$

$$= -M_{n+1}. \tag{33}$$

Now the seminvariants Λ of the function $e^{-ra}p(a)$ are defined by

$$\begin{cases} M_1 = \Lambda_1 M_0 \\ M_2 = \Lambda_1 M_1 + \Lambda_2 M_0 \\ M_3 = \Lambda_1 M_2 + 2\Lambda_2 M_1 + \Lambda_3 M_0 \\ M_4 = \Lambda_1 M_3 + 3\Lambda_2 M_2 + 3\Lambda_3 M_1 + \Lambda_4 M_0. \end{cases} \tag{34}$$

On the other hand, in view of (33)

$$\begin{cases} M_1 = \Lambda_1 M_0 \\ M_2 = \Lambda_1 M_1 - \Lambda_1' M_0 \\ M_3 = \Lambda_1 M_2 - 2\Lambda_1' M_1 + \Lambda_1'' M_0. \end{cases} \tag{35}$$

where the primes denote derivatives with respect to r.

Hence

$$\begin{cases} \Lambda_2 = -\Lambda_1' \\ \Lambda_3 = -\Lambda_2' = \Lambda_1'' \end{cases} \tag{36}$$

and generally

$$\Lambda_{n+1} = -\frac{\partial \Lambda_n}{\partial r} \tag{37}$$

that is, if successive moments are successive negative partial derivatives with respect to r, the same is true of successive seminvariants.

Furthermore, we have

$$\frac{\partial \beta(r)}{\partial r} = \beta(r) \frac{\int_0^\infty a e^{-ra} p(a)\, da}{\int_0^\infty e^{-ra} p(a)\, da} \tag{38}$$

$$= \Lambda_1 \beta(r). \tag{39}$$

Hence, in this sense, and denoting successive partial derivatives of β by subscripts

$$\begin{cases} \beta_1 = \Lambda_1 \beta_0 \\ \beta_2 = \Lambda_1 \beta_1 - \Lambda_2 \beta_0 \\ \beta_3 = \Lambda_1 \beta_2 - 2\Lambda_2 \beta_1 + \Lambda_3 \beta_0 \end{cases} \tag{40}$$

a set of equations from which successive partial derivatives of $\beta(r)$ can be obtained if the seminvariants Λ are given.

In actual computation the seminvariants Λ are calculated according to (34) from the moments M, which themselves must first be computed as functions of r. If the entire computation is required for only one particular value of r, the moments M may be calculated directly by numerical integration of (32). But if their values are required for a series of values of r, direct computation would be very laborious. Unless r is rather small, merely expanding the exponential under the integral sign and integrating term by term is unsatisfactory as the series converges too slowly. Much more rapid convergence is secured by obtaining a series development not of the moments themselves, but of the ratio between two successive moments, thus:

$$\frac{M_{n+1}}{M_n} = \frac{m_{n+1} - r m_{n+2} + \frac{r^2}{2!} \cdot m_{n+3} - \cdots}{m_n - r m_{n+1} + \frac{r^2}{2!} \cdot m_{n+2} - \cdots} \tag{41}$$

$$= \lambda_{n1} - r\lambda_{n2} + \frac{r^2}{2!} \cdot \lambda_{n3} - \cdots$$

where λ_{nj} is the jth seminvariant of $a^n p(a)$, that is,

$$(42)\qquad \begin{cases} m_n = m_n \\ m_{n+1} = \lambda_{n1} m_n \\ m_{n+2} = \lambda_{n1} m_{n+1} + \lambda_{n2} m_n \\ m_{n+3} = \lambda_{n1} m_{n+2} + 2\lambda_{n2} m_{n+1} + \lambda_{n3} m_n \end{cases}$$

Furthermore, according to (33)

$$(43)\qquad \begin{aligned} \frac{\partial M_n}{M_n \partial r} &= -\frac{M_{n+1}}{M_n} \\ &= -\left(\lambda_{n1} - r\lambda_{n2} + \frac{r^2}{2!}\lambda_{n3} - \cdots\right) \end{aligned}$$

Hence

$$(44)\qquad M_n = m_n e^{-r\lambda_{n1} + \frac{r^2}{2!}\lambda_{n2} - \cdots}$$

a formula which enables us to compute directly the moments M of $e^{-ra}p(a)$ from the moments of $p(a)$ and seminvariants of $a^n p(a)$. The seminvariants Λ and the derivatives $\beta_1, \beta_2 \cdots$ then follow according to (34) and (40).

3. **Recapitulation.** By virtue of the various properties of the moments and seminvariants thus developed, the following routine may be followed in the computation of the successive derivatives β_n. By direct computation, determine the moments m_n of $p(a)$. Then obtain in succession the several characteristics as follows, the numbers over the arrows indicating the pertinent equation in the text:

$$m_n \xrightarrow{(42)} \lambda_{nj} \xrightarrow{(44)} M_n \xrightarrow{(34)} \Lambda_n \xrightarrow{(40)} \beta_n$$

4. **Numerical example.** By way of illustration the results obtained in preceding sections were applied to a logistic population for which a series expansion of the annual births $B(t)$ in terms of the logistic and its derivatives was available from a previous computation[4] carried out by a method less general than the one here presented. Of special interest in the numerical results now to be shown is the comparison between the two representations, on the one hand $B(t)$ in terms of $\varphi(t)$, the logistic in this case, and its derivatives; on the other hand $b(t)$ the birth rate per head, in terms of $\beta(r_t)$ and its partial derivatives with respect to r.

The data on which these computations are based are derived from the actual growth of the population of the United States, which from 1790 to 1930 followed rather closely the logistic function

[4] *Human Biology*, loc. cit., *Jl. Soc. Statistique, Paris*, 1933, vol. 74, pp. 336, 341.

TABLE I

Analysis of the Birth Rate Curve in a Logistic Population Subject to a Constant Mortality

Numerical Values of Life Table and Other Characteristics, Based on the Mortality in the United States* 1919–1920

Characteristic	Defining Equation	n					
		0	1	2	3	4	5
		Moments of $p(a)$ and Seminvariants of $a^n p(a)$ According to United States Life Table for 1919–1920, White Females					
m_n	7	57.518	2014.9	99,179.	56,712. $\times 10^2$	35,407. $\times 10^4$	23,455. $\times 10^6$
λ_{n1}	12, 14	35.031	497.12	3365.	−223,390.	−74,241. $\times 10^2$	
λ_{n2}	42	49.222	391.79	−1386.	−112,242.	26,338. $\times 10^2$	
λ_{n3}	42	57.182	300.28	−1984.	−34,242.		
λ_{n4}	42	62.433	237.99	−1744.			
λ_{n5}	42	66.245	195.52				
		Moments and Seminvariants of $e^{-ra}p(a)$ and the Corresponding Values of $\left[\frac{\varphi_n}{\varphi_0}\right]_0$ and of β_n for Selected Values of r					
		Calendar Year 1800 (r = .030578)					
M_n	32	24.298	538.26	20,028.	947,340.	51,178. $\times 10^3$	2,827. $\times 10^6$
Λ_n	34		22.152	333.54	5952.0	22,238.	−13,734. $\times 10^3$
$[\varphi_n/\varphi_0]_0$	28	1.000	0	−2591. $\times 10^{-8}$	−770. $\times 10^{-9}$	−196. $\times 10^{-10}$	−515. $\times 10^{-12}$
β_n	40		.91103	6.4638	−219.77	4016.6	−32,849.

		Calendar Year 1850 (r = .027727)					
M_n	32	25.922	599.35	22,946.	11,075. $\times 10^2$	60,825. $\times 10^3$	3,444. $\times 10^6$
Λ_n	34		23.121	350.60	6,045.4	8,266.3	−14,199. $\times 10^3$
$[\varphi_n/\varphi_0]_0$	28	1.000	0	−10,256. $\times 10^{-8}$	−246. $\times 10^{-8}$	−66. $\times 10^{-10}$	634. $\times 10^{-12}$
β_n	40		.89148	7.0938	−228.00	3,116.5	72,165.
		Calendar Year 1900 (r = .019138)					
M_n	32	32.066	844.67	35,154.	17,972. $\times 10^2$	10,329. $\times 10^4$	6,239. $\times 10^6$
Λ_n	34		26.341	402.43	5,966.9	−50,322.	−14,445. $\times 10^3$
$[\varphi_n/\varphi_0]_0$	28	1.000	0	−23,515. $\times 10^{-8}$	−161. $\times 10^{-8}$	265. $\times 10^{-9}$	272. $\times 10^{-11}$
β_n	40		.82155	9.0890	−235.72	−906.36	395,810.
		Calendar Year 1950 (r = .007684)					
M_n	32	44.580	1,396.7	64,491.	35,353. $\times 10^2$	21,407. $\times 10^4$	13,783. $\times 10^6$
Λ_n	34		31.329	465.12	4,837.5	−155,820.	−11,278. $\times 10^3$
$[\varphi_n/\varphi_0]_0$	28	1.000	0	−18,242. $\times 10^{-8}$	293. $\times 10^{-8}$	119. $\times 10^{-9}$	−355. $\times 10^{-11}$
β_n	40		.70274	11.583	−182.32	−8180.6	604,140.
		Calendar Year 2000 (r = .001980)					
M_n	32	53.715	1,829.2	88,617.	50,144. $\times 10^2$	31,066. $\times 10^4$	20,452. $\times 10^6$
Λ_n	34		34.054	490.09	3790.9	−208,380.	−8,334. $\times 10^3$
$[\varphi_n/\varphi_0]_0$	28	1.000	0	−583. $\times 10^{-7}$	16. $\times 10^{-7}$	−27. $\times 10^{-9}$	482. $\times 10^{-12}$
β_n	40		.63398	12.466	−126.32	−11,541.	511,440.

* The life table for white females was used, for which the requisite constants were available from the author's prior publications.

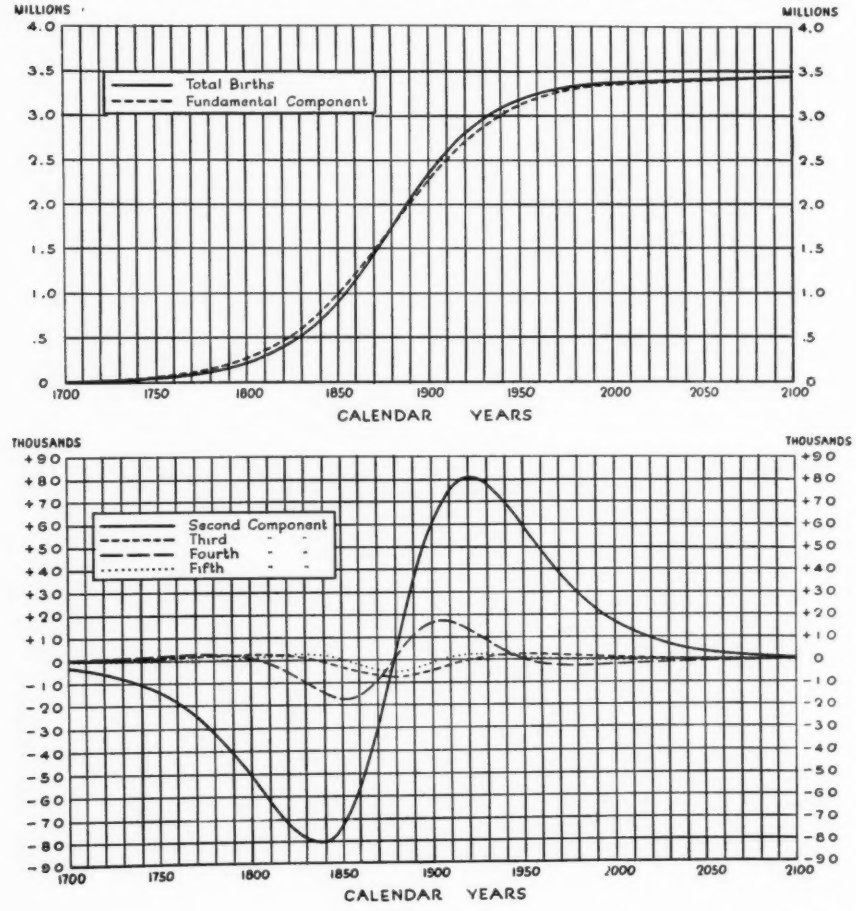

FIG. 1

$$N(t) = \frac{197{,}493{,}000}{1 + e^{-.0314(t'-1914)}} = \frac{N_\infty}{1 + e^{-.0314t}} \tag{45}$$
$$= N_\infty \Phi(t)$$

where $\Phi(t)$ is used to distinguish the special case of the logistic function, from the general case $\varphi(t)$, and where t' denotes the calendar year.

This was combined, in the computations, with the life table for white females, United States 1919–1920, supposed constant throughout the period.[5]

[5] This is, of course, an arbitrary assumption made here simply for illustrative purposes. The life table for white females was used because of related computations regarding the intrinsic rate of natural increase, which have been reported on elsewhere.

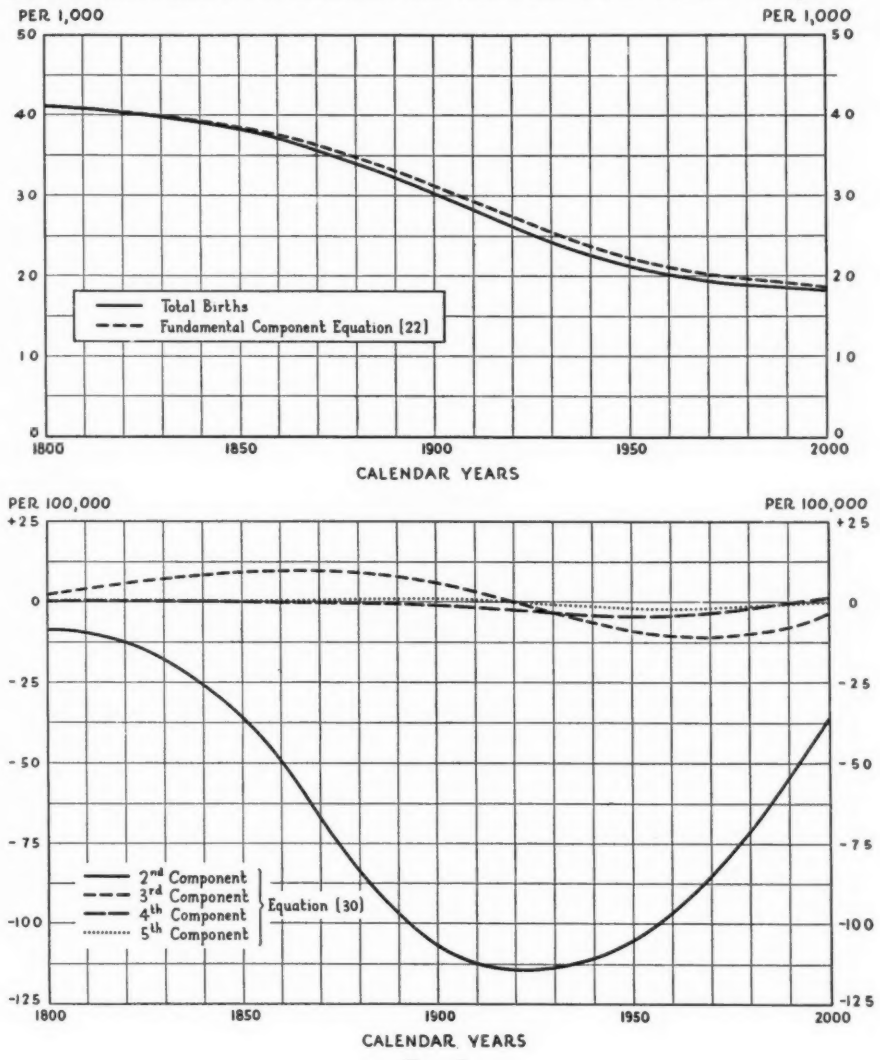

FIG. 2

With this basis, the fundamental data are as follows:

1. Quantities depending solely on the life table, namely, m_n, λ_{nj}. These are exhibited in the first section of Table I.

2. Quantities depending on the life table and also on r_t, namely, M_n, Λ_n, $\left[\frac{\varphi_n}{\varphi_0}\right]_0$, β_n. These are exhibited in the remaining sections of Table I.

5. **Comparison of the representation (20) of the annual births $B(t)$ and the representation (30) of the annual birth rate $b(t)$.** It is interesting to make this comparison, as applied to the case of the logistic population, for there are certain points of marked difference. The graphs Fig. 1 and Fig. 2 show this at a glance. In both cases the fundamental component alone yields a very fair approximation to the full solution, but the second component is of very different character in the two cases. In the composition of $B(t)$ it starts from a vanishing value, diminishes through negative values to a minimum, then, passing through zero at the "center," it rises to a maximum positive value, and finally approaches zero asymptotically from above.

The second component of $b(t)$, starting also from a vanishing value, forms a single downwardly convex loop, and then approaches zero asymptotically from below.

The higher components in both cases are relatively insignificant.

III. APPENDIX

1. **Symbols used.** It may be convenient to assemble together here certain of the symbols used in the text:

$$m_n = \int_0^\infty a^n p(a)\, da = n\text{th moment of } p(a)$$

$$M_n = \int_0^\infty a^n e^{-ra} p(a)\, da = n\text{th moment of } e^{-ra}p(a)$$

$\lambda_{nj} = j$th seminvariant of $a^n p(a)$
$\lambda_{0j} = \lambda_j = j$th seminvariant of $p(a)$
$\Lambda_j = j$th seminvariant of $e^{-ra}p(a)$

$$\beta(r) = \frac{1}{\int_0^\infty e^{-ra} p(a)\, da}$$

$$\beta_n = \frac{\partial^n \beta}{\partial r^n}$$

$$r_{1+n} = \frac{\partial^n r_1}{\partial t^n}$$

$\left[\frac{\varphi_n}{\varphi_0}\right]_0$ For definition of this, see equations (25), (26), (28)

2. **Derivatives of $\varphi(t)$ and properties of the Logistic Function.** In the particular case that $\varphi(t)$ is the logistic function $\Phi(t)$, the successive derivatives $\Phi_1, \Phi_2, \cdots$ may be obtained step by step by equations (25), (26), taking advantage of special properties of that function.

$$\text{(46)}\qquad \begin{cases} \Phi(t) + \Phi(-t) = \dfrac{1}{1+e^{-Kt}} + \dfrac{1}{1+e^{Kt}} \\ \qquad\qquad = \dfrac{1}{1+e^{-Kt}} + \dfrac{e^{-Kt}}{1+e^{-Kt}} \\ \qquad\qquad = 1 \end{cases}$$

Hence, putting

$$\text{(47)}\qquad \Psi(t) = \Phi(-t)$$

We have

$$\text{(48)}\qquad \Phi + \Psi = 1$$

$$\text{(49)}\qquad \Psi = 1 - \Phi$$

Denoting the nth derivatives by the subscript n, it follows at once from (49) that

$$\text{(50)}\qquad \Psi_n = -\Phi_n$$

$$\text{(51)}\qquad r_1 = \frac{\Phi_1}{\Phi_0} = \frac{Ke^{-Kt}}{(1+e^{-Kt})^2} \Big/ \frac{1}{1+e^{-Kt}}$$

$$= \frac{K}{1+e^{Kt}} = K\Psi_0$$

$$\text{(52)}\qquad r_{n+1} = K\Psi_n$$

Hence, in the case of the logistic, the algorithm (25) takes the form

$$\text{(53)}\qquad \Phi_1 = K\Phi_0\Psi_0 = K\Phi_0(1-\Phi_0)$$

$$\Phi_2 = K\{\Phi_0\Psi_1 + \Phi_1\Psi_0\}$$

$$= -K\Phi_1\{\Phi_0 - (1-\Phi_0)\}$$

$$\text{(54)}\qquad = K^2\Phi_0(1-\Phi_0)(1-2\Phi_0)$$

$$\Phi_3 = K\{\Phi_0\Psi_2 + 2\Phi_1\Psi_1 + \Phi_2\Psi_0\}$$

$$= K^3\Phi_0(1-\Phi_0)(1-6\Phi_0+6\Phi_0^2)$$

$$\text{(55)}\qquad = K^3\Phi_0(1-\Phi_0)\left(\frac{1}{2}+\frac{\sqrt{3}}{6}-\Phi_0\right)\left(\frac{1}{2}-\frac{\sqrt{3}}{6}-\Phi_0\right)$$

It is seen that all derivatives vanish at $\Phi_0 = 0$ and at $\Phi_0 = +1$, that is at $t = \pm\infty$. Furthermore, Φ_2 vanishes at $\Phi_0 = \frac{1}{2}$, that is at $t = 0$; and Φ_3 vanishes at $\Phi_0 = \dfrac{1}{2} \pm \dfrac{\sqrt{3}}{6}$, that is at $\tanh\dfrac{Kt}{2} = \dfrac{\sqrt{3}}{3}$, since

$$\text{(56)}\qquad \Phi_0(t) = \frac{1}{2} + \frac{1}{2}\tanh\frac{Kt}{2}$$

Successive derivatives of Φ can thus be computed successively according to (53), (54), (55), etc. For purposes of record, however, it may be convenient to note here explicit expressions for these derivatives, and a simple algorithm by which the numerical coefficients occurring in them can be written down at sight. It is found, by carrying out the differentiation directly, that

$$(57)\qquad \begin{cases} \Phi_1 = re^{rt}\cdot\dfrac{1}{(1+e^{rt})^2} \\ \Phi_2 = r^2e^{rt}\cdot\dfrac{1-e^{rt}}{(1+e^{rt})^3} \\ \Phi_3 = r^3e^{rt}\cdot\dfrac{1-4e^{rt}+e^{2rt}}{(1+e^{rt})^4} \\ \Phi_4 = r^4e^{rt}\cdot\dfrac{1-11e^{rt}+11e^{2rt}-e^{3rt}}{(1+e^{rt})^5} \\ \Phi_5 = r^5e^{rt}\cdot\dfrac{1-26e^{rt}+66e^{2rt}-26e^{3rt}+e^{4rt}}{(1+e^{rt})^6} \end{cases}$$

The numerical coefficients can be obtained by the modification of the Pascal triangle shown in Fig. 3. Its use is most easily explained by an example. Thus the coefficient -4 in the third line is obtained as the sum of the two immediately adjoining figures in the line above it, each multiplied by the rank of the oblique row in which it appears. This rank is indicated by the corresponding number written above the ruled line forming the "roof" of the triangle. Thus the second coefficient in the third horizontal line from above is obtained as $(1 \times -2) + (-1 \times 2) = -4$. Similarly, the third coefficient in the last line of the diagram (which must be regarded actually as extending indefinitely) is the sum of $(-57 \times -5) + (302 \times 3) = 1191$.

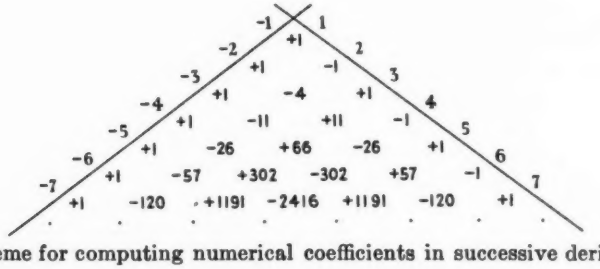

FIG. 3. Scheme for computing numerical coefficients in successive derivatives of logistic function.

2. Construction of coefficients in (26). The numerical coefficients appearing in equation (26) are constructed according to the following rules:

(a). The expression $\frac{\varphi_n}{\varphi_0}$ contains all possible products of the form $r_a r_b r_c \cdots$ the sum of whose subscripts is n, due regard being had to powers of r. Thus for example $\frac{\varphi_4}{\varphi_0}$ contains r_1^4, that is $r_1r_1r_1r_1$; also $r_1^2r_2$, r_1r_3, r_2^2 and r_4.

(b). If a, b, c are all different, the coefficient $Q_{abc\ldots}$ of $r_a r_b r_c \cdots$ is formed according to the following pattern, in which ${}^pC_q$ denotes, in the customary notation, the binomial coefficient $\binom{p}{q}$

$$Q_{abc} = {}^nC_{n-(a+b+c)}{}^{a+b+c}C_{b+c}{}^{b+c}C_c$$

$$= \frac{n!}{\{n-(a+b+c)\}!a!b!c!} \tag{58}$$

If some of the subscripts are equal, that is if some of the factors occur as the sth power of r, then the formula for Q is modified by the introduction of the corresponding factorial $s!$ in the denominator, according to the pattern of the following example:

If $b = c$, so that $r_a r_b r_c = r_a r_b^2$

then the corresponding coefficient is

$$Q_{abb} = \frac{1}{2!} \cdot {}^nC_{n-(a+2b)}{}^{a+2b}C_{2b}{}^{2b}C_b \tag{59}$$

$$= \frac{n!}{2!\{n-(a+2b)\}!a!(b!)^2} \tag{60}$$

More generally, the coefficient of $r_a^u r_b^v r_c^w \cdots$ is

$$Q_{abc}^{(uvw)} = \frac{n!}{(a!)^u(b!)^v(c!)^w \cdots u!v!w! \cdots} \tag{61}$$

where

$$a > b > c \tag{62}$$

and

$$au + bv + cw + \cdots = n \tag{63}$$

Formula (59) may be found more convenient than (60) if a table of the binomial coefficients is available; for in the case here exhibited for example, formula (59) requires only 3 tabular values to be looked up, whereas formula (60) calls for 4.

It may be noted that coefficients of this form occur in certain formulae relating to seminvariants,[6] also in the theory of partitions.[7]

METROPOLITAN LIFE INSURANCE COMPANY, NEW YORK.

[6] See for example R. Frisch, *Sur les semi-invariants et moments employés dans l'étude des distributions statistiques*, Oslo, 1926; C. C. Craig, *Metron*, 1928, vol. VII, p. 10.

[7] Dwyer, *Annals of Mathematical Statistics*, 1938, vol. 8, p. 21; vol. 9, pp. 4, 8. E. G. Olds, *Bulletin Am. Math. Soc.*, 1938, vol. 44, p. 412. H. S. Wall, *Bulletin Am. Math. Soc.*, 1938, vol. 44, p. 395; P. S. Dwyer, *Annals Math. Statistics*, 1938, vol. 9, p. 116. E. A. Cornish and R. A. Fisher, *Revue de l'Institut Internat. Statist.*, 1937, vol. 5, p. 307.

THE INTERPRETATION OF CERTAIN REGRESSION METHODS AND THEIR USE IN BIOLOGICAL AND INDUSTRIAL RESEARCH[1]

By C. Eisenhart

1. **Introduction.** Just as the scientific theorist depends upon the research worker for the facts upon which to build his theory, so does the practical man rely upon empirical relationships to help him estimate (or predict) the value of one quantity from that of another. Sometimes he is interested in assessing the value of some quantity which it is impracticable or impossible to observe directly in a given instance, the estimation being performed with the aid of a previously established relationship between the quantity whose value is sought and another whose value can be determined directly. In other instances he wishes to make use of the relationship existing between two or more quantities to help him adopt a course of action which has a good chance of leading him to a desired result. An example is that of a manufacturer who wishes to exercise control at various stages of a manufacturing process so as to produce a product whose quality lies within a specified range.

In appealing to the interests of the practical man, proponents of statistical methods have often illustrated their writings with beautiful examples of the power of this implement of research, without adequately discussing the abstract ideas that underlie the methods they have promoted—ideas essential to correct statistical thinking. The result has been that to many research workers certain problems with similar objectives appear amenable to identical statistical solution, when in fact intrinsic differences exist which alter considerably the details of their solution.

Such misinformation is particularly prevalent among those whose knowledge of the mathematics of correlation, and of curve fitting, has been derived from the treatment in elementary statistics courses of problems in which no one of the variables stands out from the rest as being *the* dependent variable, with its values determined (not exactly, but within limits) from the values that happen to be assumed by the other variables in the data under investigation. In elementary courses the usual procedure in such cases is to *take* one of the variables as the dependent variable, and then *consider* the others as independent variables. Furthermore, the curve-fitting procedure usually adopted depends on the additional assumption that the values of the independent variables are known exactly (without error)—an assumption often passed by without mention, and one that

[1] Revised from an expository paper presented, under different title, to the American Statistical Association, at Detroit, December 29, 1938, at the invitation of the program committee of the Biometrics Section.

introduces artificiality into the analysis and imposes limitations on the range of applicability of the inferences drawn. This simplification of problems without explicit mention of the fact, fosters misconceptions that are carried over into analyses of data in which the dependent variable is definitely a particular one of the variables and no other—a particularly bad misconception being that the variable whose value is to be estimated automatically assumes the rôle of the independent variable. The calculation and use of dosage-response curves in problems of biological assay constitute an example, and a case which has been correctly solved. The dosage-response curve should be evaluated from a series of observations, with dosage as the *independent* variable, and the curve then used to estimate unknown dosages from observable responses.

It is one object of the present paper to pass in review some of the ideas involved in current curve-fitting practices so that the reader can see for himself why, when one is interested in estimating X from Y, in some instances it is necessary to follow out curve-fitting practices with Y as the dependent variable, and then use the inverse of the relation found. In addition, it is an object of this paper to indicate the types of problem to which this method of inverse regression affords a solution, and to emphasize the confidence interval nature of the estimates it provides. The method will be exemplified by working out in detail a problem arising in the manufacture of cheese, and also a problem concerned with the biological assay of a hormone substance.[2]

2. **Mathematical Aspects of the Formulation Of Empirical Relationships.** Probably the most obvious way of investigating whether any relationship exists between two variables is that of plotting the observed pairs of values on graph paper. For simplicity we shall confine our attention in this paper to the case of only two variables. While the general trend of the plotted points may suggest the existence of a relationship, the plotted points themselves do not give a definite expression of that relationship, and it is often desirable to have a formula of some sort that expresses it concisely. Furthermore, in all branches of science the data of the research worker are subject to all sorts of fluctuations which tend to make the observational points scatter about a general trend in a band not unlike the Milky Way. Consequently various methods have been developed for inferring from the observations the 'true relation' between the quantities concerned, or, more exactly, a relation which it is hoped will be sufficiently close to the 'true relation' *for the purposes in mind*.

In the development of these methods two rather different viewpoints had to be taken into consideration: first, that of the physical scientist who views the irregular fluctuations as being quite apart from the phenomena under observation and arising solely from inaccuracies of measurement and experimental

[2] Those who are primarily interested in problems of biological assay will find additional material in references [26] to [31]; those whose interests lie in the direction of quality control are referred to W. A. Shewhart [9], and E. S. Pearson [5]. Numbers in [ ] refer to the references at end of the paper.

technique; secondly, that of the biological and social scientists who attribute a large portion of the apparent irregularity of their observations to a real variability which is an essential part of the phenomena studied. That two such divergent viewpoints could be brought together on a common ground is a tribute to the pioneers in mathematical statistics, and the manner in which it has been effected is indicated by the following entry in E. S. Pearson's notebook[3] for 1921–22:

"The purpose of the mathematical theory of statistics is to deal with the relationship between 2 or more variable quantities, without assuming that one is a single-valued mathematical function of the rest. The statistician does not think that a certain x will produce a single-valued y; not a causative relation but a correlation. The relationship between x and y will be somewhere within a zone and we have to work out the probability that the point (x, y) will lie in different parts of that zone. The physicist is limited and shrinks the zone into a line. Our treatment will fit all the vagueness of biology, sociology, etc. A very wide science."

When viewed from this angle, the fundamental problem in the determination of a relationship between two variables, say X and Y, is to determine as accurately as possible from the data in hand the simultaneous probability distribution of the observable quantities, say x and y, considered as random variables. There is, however, a subtle but important distinction between the cases in which the random variability of x and y is due to errors of measurement, etc., and the cases in which this random variability is, as in biological variation, a part of the phenomena under investigation. In the latter we postulate the existence of a probability distribution of the random point (x, y) about some point of location $(\bar{\bar{X}}, \bar{\bar{Y}})$, where the exact meaning of the coördinates $\bar{\bar{X}}$ and $\bar{\bar{Y}}$ depends on the nature of the probability distribution, although they will generally be the coordinates of the mode. In these cases, since (x, y) is subject to biological variation *only*, (x, y) will lie on the line $X' =$ constant only in cases where $x = X'$. Accordingly, along a line $x = X'$ we shall have the probability distribution of the random point (X', y) about some point of location $(X', \bar{\bar{Y}}_{X'})$. This may not be true when x is also or alone subject to experimental error, for here we postulate a separate probability distribution of (x, y) for each 'true point' (X, Y), and when there are 'errors' in both coördinates (x, y) can lie on the line $x = X'$ when $X \neq X'$. In these cases, the observed distribution of (x, y) for $x = X'$ may result from sampling more than one probability distribution and cannot be interpreted simply. If, however, the X-coördinate is never subject to error, the distribution of (x, y) for $x = X'$ samples the probability distribution of (x, y) for $(X', Y_{X'})$, where $Y_{X'}$ is the true value of Y for $X = X'$. Clearly similar remarks apply in terms of y and Y.

[3] E. S. Pearson, *Biometrika*, vol. XXIX, parts III and IV, (1938) p. 208, writes: "I find on page 1 of my Notes the following statement, which was probably taken down fairly closely from (Karl) Pearson's words: 'The purpose of. . . .' "

Actually at the outset it is not customary to embark on the solution of such a general problem as the determination of the simultaneous probability distribution of x and y. Instead, in the cases where both x and y are subject to 'error', it is customary to assume that the distributions of x about X, and y about Y, are of some particular functional form and then seek to estimate from the data the 'true relation' $\varphi(X, Y) = 0$. Likewise, when x and y are subject only to biological variation, say, it is customary to seek an estimate of the functional relation $\varphi(X, \bar{Y}_X) = 0$, or of the relation $\varphi(\bar{X}_Y, Y) = 0$, where $\bar{Y}_X$ and $\bar{X}_Y$ denote some sort of average (not necessarily the mathematical expectation) of y for a given X and of x for a given Y, respectively, the former being interpreted as being the 'true relation' between X and the average value of y for that X, with a similar interpretation for the latter function. Furthermore, in these cases of mere biological variation it is customary to take $x \equiv X$, $y \equiv Y$, that is, to assume that what we observe are the true values of the quantities, that any errors of measurement are negligible compared to the sampling fluctuations arising from real biological variation.

So far as I know all methods of utilizing observed values of two variables to obtain a relation between the two variables that it is hoped will be sufficiently close to the true relation for the purposes in mind involve the following steps:

(1) To assume that the observational points $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$ differ from the points $(X_1, Y_1), (X_2, Y_2), \cdots, (X_N, Y_N)$ as the result of observational errors[4] involved in the x, or in the y, or in both coördinates.

(2) To assume, either from the general appearance of the graph of the plotted points or from theoretical considerations, that the relationship between X and Y *is* of the form $\varphi(X, Y; \alpha_0, \alpha_1, \cdots, \alpha_{k-1}) = 0$, where φ is some definite mathematical function involving k, $k \leq N$, constants whose values are unknown. If it is not assumed that φ *is* the true functional relation between X and Y, then it is assumed that the functional relation specified by the φ will be adequate for the purposes in mind.

(3) To choose as an estimate of φ the function $\hat{\varphi} = \varphi(X, Y; a_0, a_1, \cdots, a_{k-1})$ where the a's are those values of the α's that render $\hat{\varphi}$ the function of form φ which is the best fit to the observed points (x_i, y_i), $(i = 1, 2, \cdots, N)$, in some sense of the word "best"; and finally, a step which is too often overlooked.

(4) To carry out some test of goodness of the fit of $\hat{\varphi}$ to the observed points upon the outcome of which rests the decision as to whether a function of the form φ can adequately describe the observed relation between the x's and y's, and, if the decision be affirmative, accepting $\hat{\varphi}$ as an estimate of the true function of form φ.

---

[4] The word "error" here should be interpreted as "experimental or technical error" from the viewpoint of the physical sciences, (which errors are unbiased in the sense that they average out in the long run), and as "biological variation" from the viewpoint of the biological scientist. In the latter case, if the biological variation is involved in y, and not in x, then $x_i \equiv X_i$ and $Y_i \equiv \bar{Y}_{X_i}$; a similar statement holding if x is in error but not y. In the former case, X and Y are the "true values" of the variables.

In connection with step (4) some results were obtained by W. E. Deming [2] for the case where φ is fit by the method of least squares. He has found that the sum of squared residuals, which is the function to be minimized by the fitting procedure, is fairly sensitive to changes in the functional form of φ, that is, to changes which alter its graph within the range of the observations, but much less sensitive to changes in the values of the parameters involved in a particular functional form. Consequently, by comparing the minimum values of the sums of squared residuals for two different functional forms φ_1 and φ_2 under tentative consideration, it will often be possible to make a good choice between them. On the other hand, it may be possible to alter considerably the values of the parameters in the functional form chosen without appreciably altering the value of the sum of squared residuals. From this it is seen that φ may not be well determined by $\hat{\varphi}$ even though the functional form of φ may be the correct one for the relationship under investigation. For the case where X is exactly known for each observation, with only y subject to error, Deming shows that for the same sum of squared residuals φ is better determined by $\hat{\varphi}$ when there is a long range in X than when there is a short range. In terms of the measure of goodness of fit appropriate to any method of curve fitting these conclusions will probably carry over to that method of curve fitting.

Step (2) also deserves further comment: The function φ may be such that $\hat{\varphi}$ fits the data well within the range of x and y studied, but it must be remembered that an infinite number of other formulae exist which could be adjusted so as to fit the observed points equally well, and some might be found which could be made to fit better. Once a particular functional form for φ has been chosen, if $\hat{\varphi}$ is used to "extrapolate" beyond the range of the observed points, or, if $\hat{\varphi}$ is used as *the* relation between X and Y in any theoretical considerations, it must be remembered that the soundness of any inference that can be made rests to a large extent on the validity of the logic or theoretical considerations that lead to the choice of φ as the expression of the functional relation between the variables, and that the goodness of fit of $\hat{\varphi}$ for one particular batch of data is not a justification of these extensions.

3. Some general remarks on curve fitting practices.

In many cases the assumption is made that a linear relation prevails between X and Y, that is, it is assumed that

$$\alpha_0 + \alpha_1 X + \alpha_2 Y = 0 \tag{1}$$

which may be written in the equivalent forms

$$Y = \alpha + \beta X, \quad \text{where } \alpha = -\alpha_0/\alpha_2, \quad \text{and } \beta = -\alpha_1/\alpha_2 \tag{2}$$

$$X = \gamma + \delta Y, \quad \text{where } \gamma = -\alpha_0/\alpha_1, \quad \text{and } \delta = -\alpha_2/\alpha_1. \tag{3}$$

We are adopting for the moment the viewpoint of the physical scientist, and assuming that (1) represents the true relation between X and Y. We shall return to the case of biological variation later.

A common impression on the part of the research worker, regarding the principles of curve-fitting, seems to be: If one is interested in estimating Y from X, then take $\hat{Y} = a + bX$ as the estimate of (2), and therefore of (1), the a and b being those values which make the line a good fit in terms of the deviations $(y - \hat{Y})$—if one were fitting by the method of least squares one would find the a and b that minimize $\Sigma(y - \hat{Y})^2$, Σ denoting summation over the observed values of y and their corresponding $\hat{Y}$ values; on the other hand, if one is interested in estimating X from Y, then $\hat{X} = b + cY$ is to be fitted, the values of c and d being chosen so as to make $\hat{X}$ a good fit in terms of the deviations $(x - \hat{X})$. *It does not seem to be generally realized that the fitting should be done in terms of the deviations which actually represent "error." Thus when the research worker selects the X values in advance, and holds x to these values without error, and then observes the corresponding y values, the errors are in the y values, so that even if he is interested in using observed values of Y to estimate X, he should nevertheless fit $\hat{Y} = a + bX$ and then use the inverse of this relation to estimate X, i.e. $X = (\hat{Y} - a)/b$, with the best available estimate of Y substituted for $\hat{Y}$.* The situation is quite clear if one approaches the problem from the point of view of fitting the formula to the data with proper attention to which of the variables is in error, as has been recognized for a long time by writers on least squares. If both variables are in error, then this approach also leads to the appropriate solution.[5]

In order to clarify this point it will be helpful to examine the matter a little closer from the viewpoint of the *theory* of least squares.

Let us consider the case where the values of X are selected (or adjusted) by the research worker and then the corresponding values of Y found by observation. So far as the *method* of least squares is concerned in any given instance one *could* minimize $\Sigma(y - \hat{Y})^2$ *and* $\Sigma(x - \hat{X})^2$, thereby obtaining the two lines

$$\hat{Y} = a + bX \tag{4}$$

$$\hat{X} = c + dY, \text{ respectively,} \tag{5}$$

and, unless there existed a perfect correlation between the observed values of X and Y—i.e. unless all of the observed points were exactly collinear, these two fitted lines would differ and yield different estimates of (1). There is nothing in the *method* of least squares to help us choose between these, but from the viewpoint of the *theory* of least squares the correct choice in a given instance is quite clear.[5] The results of the two fitting processes may be given side by side as follows:

---

[5] See, for example, Deming [1]. Deming pays his respects to a paper by Kummel in *The Analyst* (Des Moines) vol. 6 (1879), pp. 97–105; also to a paper by Uhler, J. Optical Soc. vol. vii (1923), pp. 1043–1066.

(6)

$\Sigma(y - \hat{Y})^2$ minimized		$\Sigma(x - \hat{X})^2$ minimised	
$b = \dfrac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2}$		$d = \dfrac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(y - \bar{y})^2}$	
$a = \bar{y} - b\bar{x}$		$c = \bar{x} - d\bar{y}$	
Analysis of Variance I	*d.f.*	*Analysis of Variance II*	*d.f.*
Total variability of y's about their mean: $\Sigma(y - \bar{y})^2$	$N - 1$	Total variability of x's about their mean: $\Sigma(x - \bar{x})^2$	$N - 1$
Reduction effected by (4): $b\Sigma(x - \bar{x})(y - \bar{y})$	1	Reduction effected by (5): $d\Sigma(x - \bar{x})(y - \bar{y})$	1
Deviations about $\hat{Y}$: $\Sigma(y - \bar{y})^2 - b\Sigma(x - \bar{x})(y - \bar{y})$ $= \Sigma(y - \hat{Y})^2$	$N - 2$	Deviations about $\hat{X}$: $\Sigma(x - \bar{x})^2 - d\Sigma(x - \bar{x})\ (y - \bar{y})$ $= \Sigma(x - \hat{X})^2$	$N - 2$

In all instances Σ denotes summation of the expression following it over all the observed values; $\bar{x} = (1/N)\Sigma x$, the arithmetic mean of the *chosen* values of X; and $\bar{y} = (1/N)\Sigma y$ the arithmetic mean of the *observed* values of Y. The expression in the middle row of each table of the analysis of variance is an immediate consequence of the minimizing process employed; the last row is obtained by subtraction.

Let us now interpret these analysis of variance tables. *On the left,* $\Sigma(y - \bar{y})^2$ gives a measure of the observed variability of the y values, a portion of this variability being due, we suppose, to the dependence of Y on X. The second row of table I gives the portion (the maximum portion on the basis of the observations) of the observed variability of the y's that can be attributed to the dependence of Y on X, and the last row indicates the magnitude of the rest, that is, the magnitude of the portion of $\Sigma(y - \bar{y})^2$ that must be attributed to "error" (and, this portion has been minimized by the fitting process). *In short, remembering that we are dealing with the case in which the values of X are chosen by the research worker and only the values of Y are subject to error, the relation between X and Y being as in* (1) *or its equivalent form* (2), *we see that the analysis of variance table on the left separates $\Sigma(y - \bar{y})^2$ into portions whose meanings are clear.* In particular, since unrelated variables can exhibit relationship in finite samples, the test of whether β is really different from zero resolves itself into examining whether the variance ratio

$$\left(\frac{b\Sigma(x - \bar{x})(y - \bar{y})}{1}\right) \Big/ \left(\frac{\Sigma(y - \bar{y})^2 - b\Sigma(x - \bar{x})(y - \bar{y})}{N - 2}\right)$$

is of a magnitude that *may* be taken to indicate $\beta \neq 0$ in the sense that the risk of falsely rejecting the hypothesis that $\beta = 0$ by so doing is of an acceptable smallness.

The analysis of variance table *on the right*, on the other hand, can be misleading if it is interpreted hastily. In the first place $\Sigma(x - \bar{x})^2$ represents the variability in the chosen values of X which resulted from the way in which the research worker selected (or adjusted) them, and it is to be noted that the corresponding values observed for Y have in no way entered into their determination. Consequently the *apparent* dependence of the x on the y, measured by d, or more effectively by the second row of table II, is a spurious dependence, and the last row of this table cannot be interpreted as being a measure of the "error" in the x values, in the sense of being that portion of the variability of the x values which cannot be accounted for by the variability of the y values. *Briefly stated, when the values of x have been selected by the research worker and the corresponding y values observed, the line obtained by minimizing* $\Sigma(x - \hat{Y})^2$ *is meaningless, and* (4) *is accordingly the only correct estimate of the postulated linear relationship between X and Y, wherefore, if it is desired to reason from Y to X this must be done by means of* $X = (\hat{Y} - a)/b$, *namely* (4) *solved for X.*

In the preceding paragraphs we have discussed the case where one of the variables is subject to random variation, and the other takes only those values selected (or, to which it is adjusted) by the research worker. Without loss of generality we took Y to be the former variable, and X the latter. Actually we have discussed only the case in which (1), or one of its forms, (2) or (3), is assumed to express the 'true relation' between X and Y. That is, we have been discussing the case where y varies about Y as a result of experimental 'error,' and we have not treated the case where y is subject to biological variation.

If X takes only those values selected by the research worker, and y is subject to biological variation but is known without observational error, so that $y = Y$, (1) no longer applies for the reasons given in section 2, but it must be replaced by

$$\alpha_0 + \alpha_1 X + \alpha_2 \bar{Y}_X = 0 \tag{7}$$

where $\bar{Y}_X$ is the 'average' value (but not necessarily the arithmetic mean or mathematical expectation) of Y for the value of X denoted by the subscript. Clearly (7) may also be written in a form corresponding to (2),

$$\bar{Y}_X = \alpha + \beta X, \quad \text{with} \quad \alpha = -\alpha_0/\alpha_2 \quad \text{and} \quad \beta = -\alpha_1/\alpha_2 \tag{8}$$

or in a form corresponding to (3),

$$X = (\bar{Y}_X - \alpha)/\beta = -\alpha/\beta + (1/\beta)\bar{Y}_X . \tag{9}$$

With this latter form we may contrast

$$\bar{X}_Y = \gamma + \delta Y \tag{10}$$

a relation expressing "the true average value of X for a given Y" as a linear function of Y. Equation (10) is of interest, as well as (8), when X is free to vary in samples according to the biological variation associated with it, but when the distribution of values of X is dictated by the wishes of the research worker, as in the case under discussion, it can be demonstrated that (10) is of no value for purposes of inference.

The method adopted for estimating (7), or one of its alternative forms, will depend on what "average" $\bar{Y}_x$ is taken to be. If, as is usually the case, $\bar{Y}_x$ denotes the true arithmetic mean (or mathematical expectation) of Y for a given value of X, then (4) fitted by the method of least squares as above affords an unbiased estimate of (8). Or, if $\bar{Y}_x$ were taken to be the true median of Y for a given X, then in general one would fit (4) by minimizing $\Sigma \mid y - \hat{Y} \mid$, the summation being taken over the observed y values. As in the discussion of the case involving experimental error, to estimate X from Y one would estimate (9) with (4) solved for X, and in a particular instance replace $\hat{Y}$ by the best available estimate of $\bar{Y}_x$ from the data in hand. This brings out the strong similarity between statistical procedures appropriate when the variables are subject to experimental error and when on the other hand they are subject to biological variation but can be accurately observed.

A great injustice would be done to many previous writers by failure to mention at this point that the ideas and the conclusions reached in the preceding paragraphs have been appreciated for a long time by some of the writers who have developed the theory and applications of curve fitting. At most, the preceding paragraphs are but an emphatic way of presenting what these experts would regard as obvious.

4. **Effect of Limiting the Range of Either Variable in the Sampling Process.** In the preceding section we have discussed the situation in which one of the variables does not vary at random, but assumes only those values selected by the research workers. We have seen that in such cases this variable must be taken as the independent variable in applying any curve-fitting procedures. The same conclusion applies when both of the variables are subject to biological variation but the sampling process limits the observed range of one of the variables—only the results obtained by using the restricted variable as independent variable can be expected to give an unbiased description of the underlying relationship in the population sampled. If X is the variable for which the range of observable values is constricted by the sampling process, this means that the relation (8), for the population sampled, can be estimated from the data; but relation (10) for the population is unattainable.

To illustrate this point it will be sufficient for our purposes to consider Figure 1 which has been constructed from some artificial data which are especially suited to this purpose. We shall suppose that Y is the dependent variable and X the independent variable, and that the complete array of points shown arose from a sampling process in which neither X nor Y was restricted. It will be noticed that the observational points lie in a band sloping upward to the right and that as x increases by one unit the distribution of the corresponding y's moves up by one-half a unit. We may consider the points of the entire band shown as portraying the relationship between X and Y in the large, that is, when a point (x, y) is selected at random without restrictions on either X or Y. The slanting line labelled (I) indicates the "average" relationship prevailing between Y and

X, that is, for a given value of X the arithmetic mean of the corresponding observed values of Y is given by the point on this line with abscissa X.

Let us now consider the situation in which the points have been selected with restriction on X. As the results of such a procedure of selection let us take only those points between the two vertical lines drawn just to the right of $X = 3$

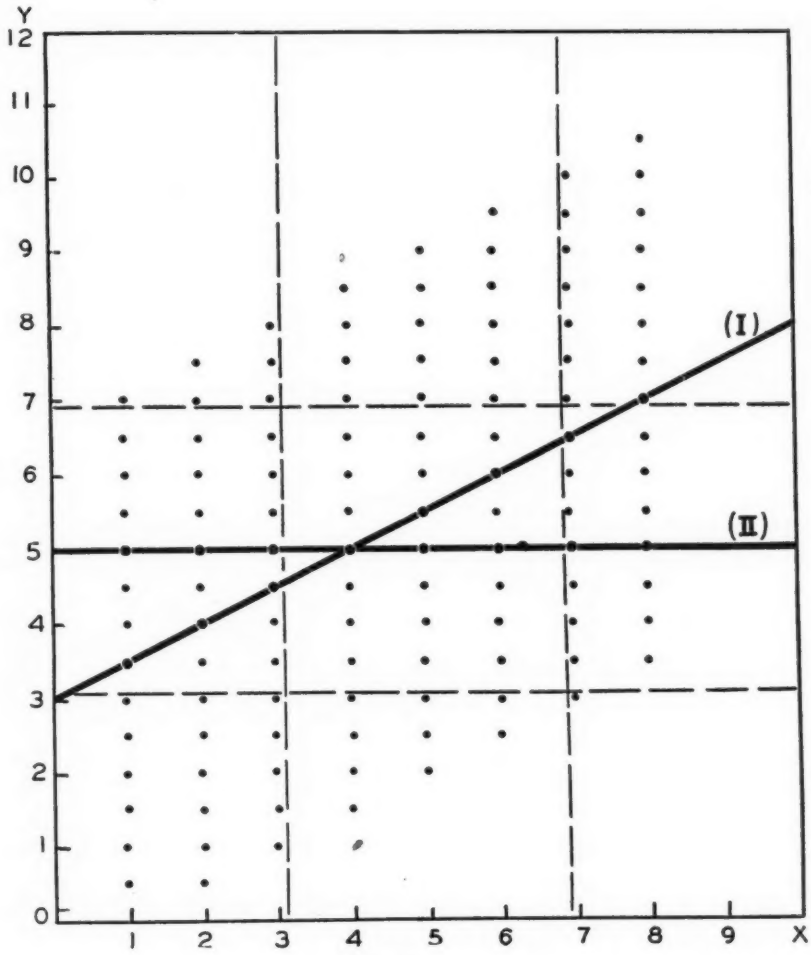

FIG. 1

and just to the left of $X = 7$. It will be seen that this does not upset the average y for a given value of x within the prescribed limits, i.e. $\bar{Y}_x$ is unaltered for $3 < X < 7$. *In other words, the introduction of a restriction with regard to* X, the independent variable, *has not spoiled the inferences with regard to* Y, *when* Y *is considered as the dependent variable*—that is, when we are arguing from X to Y.

Consider now the effect of restricting the observed y in a sampling process

and then trying to infer about $\bar{Y}_x$ in the population at large from given values of X. In Figure 1 this corresponds to considering, say, only those points that lie between the horizontal lines just above $Y = 3$ and just below $Y = 7$. It is seen immediately that in this case, i.e., between the *horizontal* lines, for every value of X the average of the observed Y values is $Y = 5$, and consequently the relation of Y to X is portrayed by the line numbered (II). *It is seen that in this case the "apparent" relation is not the correct one. Accordingly, we conclude that the restriction of the dependent variable is liable to seriously distort the relationship, so that what is observed is not representative of the true underlying situation.*

The demonstration that we have chosen is simple and artificial but the conclusions drawn apply in general, namely, the restriction of X does not alter the regression of Y on X, but the restriction of Y does. For further illustrations and a very readable discussion see Chapter 19 of *Methods of Correlation Analysis* by Mordecai Ezekiel.

As a special case of a situation in which the "observed" y's are restricted in some way or other we may turn the problem around and note the limiting case where Y is not a random variable at all but is given certain assigned values by the research worker and the corresponding values of X are ascertained by observation. It is evident from what has gone before that in such a case any formula that expresses the average value of y for a given value of x for the data thus collected is useless for inferring anything about the average value of Y for a given value of X in the "population" at large.

5. **Variables Subject to Biological Variation and also to Errors of Observation.** In the preceding paragraphs we have been supposing that the variables were subject either to errors of measurement, or to biological variation, but we excluded the case in which both types of variation were in operation simultaneously. It is reasonable to suppose that errors of measurement are present in biological work just as they are in the physical sciences, though it will usually be found that the variability between biological specimens is far greater than the maximum variability that could be attributed to errors of measurement. Accordingly, in most biological work true biological variations force errors of measurement into the background. It is usually possible to check up on this by making two or more determinations for each specimen and then comparing the variation between determinations with the variation between specimens by means of the analysis of variance technique developed by R. A. Fisher [3]. When only one determination is made per specimen the two variations cannot be distinguished.

Even if observational errors are in the background, it is of importance to know the consequences to be expected when they are superimposed on biological variation. Ezekiel discusses this phase of the subject in detail in chapter 19 of his book mentioned earlier, and a survey of his conclusions in terms of what we have discussed above will be sufficient for our purposes: (*a*) If $\bar{Y}_x$ denotes the

average value of Y corresponding to the X denoted by the subscript (in a certain sense of the word "average") and is a linear function (8) of X, then if the X values are free from errors of measurement but the y values are subject to random errors, uncorrelated with the true Y values, and which average out in the long run (in the same sense of "average" as above), then (4) fitted by the method consistent with the meaning of "average" provides an unbiased estimate of (8), in the sense that its "average" value in repeated sampling will be (8), and the effect of the errors of measurement is merely to decrease the precision with which (8) can be estimated from the given set of X values; (*b*) if the situation is as in (*a*) with the exception that the errors are correlated with the true Y values, then not only will their presence affect the precision of (4) as an estimate of (8), but it will render (4) a biased estimate of (8), the tendency being an underestimation of the existing correlation; (*c*) if random errors affect the independent variable correlated or uncorrelated with its true values, then (4) will be an unreliable estimate of (8), and may be markedly biased whether or not the errors of measurement affect the dependent variable; and, if non-random errors of measurement are present they tend to render (4) a more or less unreliable estimate of (8), quite regardless of the variables to which they apply.

The practical significance of these principles in regard to variables subject to biological variations is that if large errors of measurement enter into the determination of some variable, provided these errors are *random* that variable may still be used as the dependent variable without introducing appreciable bias in the estimation equation if enough observations are available to approximately balance out the errors; but any use of that variable as the independent variable will almost surely yield results that understate the actual relationship, and if the errors are not random, they will tend to bias the results quite regardless of the variables affected by them.

6. **An Industrial Problem.** With the preceding discussion in mind let us now direct our attention to a problem that arises in connection with the manufacture of cheese. One of the measures of the quality of a cheese is the percent of fat it contains. In the cheesemaker's notation this is given by the fat-drymatter ratio, F/DM, which is usually written as percent since the fat is contained in the total dry matter. Experience in cheese making has shown that the casein-fat ratio, C/F, of the milk out of which the cheese is made influences the F/DM of the finished cheese, and that the relationship is approximately linear, with a negative slope, for the range of values of these variables usually studied.

Since 45% is the lower limit of F/DM for an acceptable cheese as specified by law, cheese manufacturers are interested in standardizing the C/F ratio of the milk they use, which they can do by separating the milk and cream from individual sources and then putting them together again in proper proportions so that the resulting cheese will have a good chance of meeting the legal requirement *at least*. Figure 2 portrays some results obtained by standardizing the C/F ratio at different values, the individual points representing 149 different

batches of cheese manufactured in October, 1936 at a particular factory.[6] It is seen that the relationship prevailing between C/F and F/DM in these data takes the form of a rather wide sloping band and not as a close clustering of points about a well-defined trend.

If a cheese manufacturer is able to infer from data of this sort a reliable answer to a question like the following, he will be able to improve the economic efficiency of his plant: "To what value should C/F be standardized in order that we may expect F/DM to *exceed* 45 in, say, 95% of our future experience?" Unfortunately this type of question, very easy to phrase, is usually exceedingly difficult to answer, and, indeed, the very existence of an answer depends on an assump-

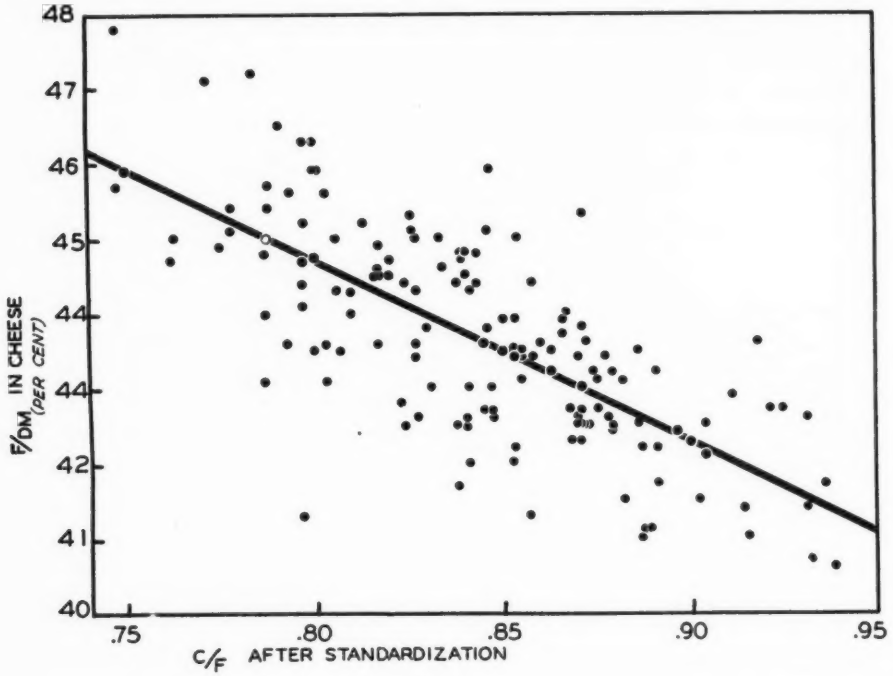

FIG. 2

tion of some sort of stability in the manufacturing process, and in the materials used, which enables a future observation to be estimated at least within limits

[6] These data were brought to me by Professor Walter V. Price, of the Department of Dairy Industry of the University of Wisconsin, in connection with a different but related problem, and I wish to acknowledge my gratitude to him for permission to use them in the present discussion. It will be noted that F/DM is given as a per cent, whereas C/F is given as a decimal fraction. This is the customary procedure with dairymen, and arises from the fact that C/F is merely an index involving two different quantities distinguishable in the milk, and cannot be interpreted as a per cent in the same way as the F/DM ratio.

from available experience. In the succeeding paragraphs we shall present a solution that will depend for its applicability upon the following assumptions:

Let Y denote the true F/DM ratio of a finished cheese, X the true C/F ratio in the milk from which it was made, and let $\bar{Y}_X$ denote the true arithmetic mean of Y associated with the value of X indicated by the subscript.

Assumption I: We shall assume that the dependence of $\bar{Y}_X$ on X is linear and given by

$$(8') \qquad \bar{Y}_X = \alpha + \beta X = \alpha' + \beta(X - \bar{x}), \quad \text{with} \quad \alpha' = \alpha + \beta\bar{x}$$

where $\bar{x}$ denotes the arithmetic mean of the true C/F values corresponding to the points shown in Figure 2.

It should be noted that $\bar{x}$ and its value do not enter into the specification of the linear relationship but only into the alternative expression of it.

Assumption II: We shall assume that X is determined without error in a given instance, and the differences $(y_X - \bar{Y}_X)$ between the observed values of F/DM, say y_X, and their corresponding mean values, $\bar{Y}_X$, may be regarded as drawn independently at random from a population in which $(y_X - \bar{Y}_X)$ are normally distributed about zero with a variance, $\sigma^2_{Y \cdot X}$, which is the same for all values of X.

Since these assumptions are restrictive it is necessary in connection with a given practical problem to ascertain whether they are acceptable on the available evidence before proceeding to the application to the problem in hand of methods depending on them for validity. Before applying to a problem of his own any of the methods presented in the following paragraphs, the reader should investigate the tenability of these assumptions with regard to his type of data. Methods for examining whether data of a given type exhibit "statistical control" are available in the literature and the reader is referred especially to the writings of W. A. Shewhart [9, 10]. To date experience has shown that it is very difficult to attain and maintain statistical stability in connection with industrial processes. On the other hand, it is uselsss to try to answer questions of inference such as the above until a fair degree of statistical stability is attained, whether statistical processes are employed or not. The success along these lines that has been attained in industry is a great tribute to Shewhart and his insistence on attention to this phase of the application of statistical methods to practical problems. The sooner workers in other fields turn their attention to questions of statistical control, the sooner mathematical statistics will be of some value to them.

From an examination of C/F and F/DM values from the same factory over a period of months it appears that although a relation of the type (8′) above seems to exist in most instances, it is not stable with regard to the values of α and β. Consequently, unless the source of this instability can be discovered and either removed, or allowed for, the answer to the above question is more or less unattainable. In order to exemplify the method, however, we shall proceed as if statistical control were a fact and assumptions I and II tenable.

It is clear, I think, from comments in the early part of this paper that if we let $Y = F/DM$ and $X = C/F$, since the C/F values have been chosen by the cheese makers, we shall have to infer about X from the relation of Y to X, the latter being considered as the *independent variable.* Furthermore, it is a consequence of assumption II that fitting

$$\hat{Y} = a + bX = a' + b(X - \bar{x}) \tag{4'}$$

by least squares will provide the most accurate estimates of α and β in (8'). That $a' = \bar{y}$, the arithmetic mean of the observed y values is evident when (4') is compared with (4) and (6). Performing the calculations it was found that

$$\hat{Y} = 64.38 - 24.58X = 43.63 - 24.58\,(X - .8439), \tag{11}$$

for the data shown in figure 2.

If now we ask "What value of C/F will to the best of our knowledge result in $F/DM = 45$ *on the average* in the future?", the answer is obtained by setting $\hat{Y} = 45$ in (11), solving for X, from which it is found that $C/F\ (= X)$ should be taken equal to $(64.38 - 45.00)/24.58 = .7884$, and this point is indicated by the black dot with white center on the line in Figure 2. We must remember, however, that (11) is merely *an estimate* of (8'), and that the value of $\hat{Y}$, namely 45, obtained by inserting $X = .7884$ in (11), is merely an estimate of the true $\bar{Y}_{.7884}$, which may not be 45 at all. Indeed the use of $\hat{Y}$ for a particular value of X to estimate the true $\bar{Y}_X$ for that X is mathematically equivalent to the customary procedure of using $\bar{y}$, the mean of all of the observed y to estimate $\bar{Y}$, the true mean of the Y population.

In recent years it has become customary to perform such estimations, not by single value, but by means of confidence intervals, a confidence interval for $\bar{Y}$ being of the form

$$\bar{Y}_1 \leq \bar{Y} \leq \bar{Y}_2$$

where $\bar{Y}_1$ and $\bar{Y}_2$ are functions of the observed values of Y, i.e. of $y_1, y_2, \cdots y_N$, and of the confidence coefficient chosen. If a confidence coefficient, of $1 - \epsilon$ is adopted $(\epsilon > 0)$, then the interpretation of such an inequality is as follows: If inequalities of this form are used whenever it is desired to estimate $\bar{Y}$ from the observed y's, then in the long run we may expect $100 \cdot (1 - \epsilon)\%$ of such estimations to be correct, that is, in $100 \cdot (1 - \epsilon)\%$ of the cases in which we apply intervals of form (6) they will include $\bar{Y}$ within their limits. Such limits are sometimes referred to as *fiducial limits* and the associated degree of confidence termed the *fiducial probability* of the estimation being correct.[7]

---

[7] There is an ever-growing literature on this mode of estimation, and a list of references to expository treatments of the subject will be found at the end of the paper together with a few other pertinent references.

From Fisher's 1935 paper it appears that he wishes to restrict the use of the words *fiducial probability, fiducial limits,* etc. to the cases in which a sufficient statistic exists for the parameter to be estimated. Since he introduced the use of these words in this con-

We shall now show how to set up confidence intervals for $\bar{Y}_X$ in terms of $\hat{Y}$ for that X, and by an extension of the argument, we shall show how to make a probability statement about the difference $(y' - \hat{Y})$ in repeated sampling, where y' is an observation not involved in the evaluation of $\hat{Y}$. The connection of this type of probability statement to the question asked above will be pointed out and its relation to the ideal answer to that question discussed.

In the succeeding paragraphs we shall make use of the following mathematical results:

(A) Assumptions I and II imply that in repeated samples *involving the same values of* X the fitted line $\hat{Y}$ of (4′) will be normally distributed about the true line $\hat{Y}_X$ of (8′) with a variance

$$\sigma^2_{\hat{Y}} = \sigma^2_a + (X - \bar{x})^2\sigma^2_b \tag{12}$$

in which

$$\begin{aligned} \sigma^2_a &= \sigma^2_{Y\cdot X}/N \\ \sigma^2_b &= \sigma^2_{Y\cdot X}/\Sigma(X - \bar{x})^2 \end{aligned} \tag{13}$$

where Σ denotes summation over the N actual values of X involved, $\bar{x}$ is the arithmetic mean of these values of X, and $\sigma^2_{Y\cdot X}$ is the true variance of Y for a fixed value of X (and assumed independent of X). The condition that the sampling be confined to the same values of X is an essential part of the statement as can be seen from the original argument by Working and Hotelling [12] which is outlined by Rider [6]. The result is given by Fisher [3] sec. 26.

(B) When $\sigma^2_{Y\cdot X}$ is unknown, a convenient estimate from the sample is

$$s^2_{y\cdot x} = \Sigma(y - \hat{Y})^2/(N - 2), \tag{14}$$

the distribution of $(N - 2)\, s^2_{y\cdot x}/\sigma^2_{Y\cdot X}$ being as χ^2 with $N - 2$ degrees of freedom and independent of the distribution of $(\hat{Y} - \bar{Y}_X)$.[13]

(C) *Student-Fisher theorem*: The ratio of any quantity d normally distributed about zero with standard deviation σ, to an estimate s having the property that ns^2/σ^2 is distributed *independently* of d as χ^2 with n degrees of freedom, is itself distributed as Student's t for n degrees of freedom.[8]

Letting $S^2_{\hat{Y}}$ denote the estimate of $\sigma^2_{\hat{Y}}$ obtained by substituting $s^2_{y\cdot x}$ for $\sigma^2_{Y\cdot X}$ in the quantities (13), it follows from (A)–(C) that

$$t = \frac{\hat{Y} - \bar{Y}_X}{S_{\hat{Y}}} \tag{15}$$

---

nection, he has some sort of right to specify their usage. Accordingly Neyman's confidence intervals are of more general availability, and when a sufficient statistic does exist both the fiducial limits and the limits of Neyman's shortest confidence limits (or of his short unbiased confidence intervals) will be found to depend on this sufficient statistic, although the interval between the limits may not be the same in the two cases, Neyman bringing an additional principle into play to assist in the location of his intervals.

[8] Fisher [4]; "Student" [11].

is distributed as Student's t for $N - 2$ degrees of freedom. Consequently if $t_{.05}$ denotes the number for which $P\{|t| > t_{.05}\} = .05$ where t is as in (15), and $|t|$ denotes the numerical value of t, it follows that the probability is .95 that random variations in the y's for the values of X chosen will yield a value of $\hat{Y}$ for which

$$-t_{.05}S_{\hat{Y}} \leq \hat{Y} - \bar{Y}_X \leq +t_{.05}S_{\hat{Y}} \tag{16}$$

is true, that is, a value of $\hat{Y}$ for which

$$\hat{Y} - t_{.05}S_{\hat{Y}} \leq \bar{Y}_X \leq \hat{Y} + t_{.05}S_{\hat{Y}} \tag{17}$$

is true. Accordingly, if we assert *in a given instance* that (17) is true, *there is no way of telling whether our assertion is correct*, but in the long run the $\hat{Y}$'s we calculate from the data we observe may be expected to differ from their $\bar{Y}_X$ values in such manner that (16) will be correct in 95% of our experience, so that we may expect to be correct in 95% of the assertions we make about $\bar{Y}_X$ using (17).

For the data of figure 2 the quantities needed in addition to (11) are

$$\frac{1}{N} = \frac{1}{149} = .00671141 \qquad \Sigma(X - \bar{x})^2 = .274796$$

$$s^2_{y\cdot x} = .9448 \qquad t_{.05} = 1.979, \text{ for 147 degrees of freedom.}$$

For $X = .7884$ it is easy to verify that $(X - .8439)^2 = .0030$, and substituting in (12) with $\sigma^2_{Y\cdot X}$ replaced by $s^2_{y\cdot x}$ gives $S_{\hat{Y}} = .1290$ *for* $X = .7884$, and, since $\hat{Y}$ equals 45 for this value of X, we may assert

$$44.744 \leq \bar{Y}_{.7884} \leq 45.256, \tag{18}$$

and we are correct in this assertion unless a 1 in 20 chance event has occurred. Since these limits do not differ widely from 45, we see that we *may hazard* the prediction that, if $X = C/F$ is standardized to .7884, then the values of $Y = F/DM$ in our future experience will be distributed about a mean fairly close to 45. This prediction is based not only on the assumption that we are sampling a stable statistical population, but also on the presumption that (18) *is* true. $\bar{Y}_{.7884}$ may really lie outside and at quite a distance from this interval. The results of a sampling experiment which illustrate this point in connection with confidence limits for a sample mean will be found in Shewhart [10].

Let us now see how the preceding type of argument may be extended to take into consideration a single additional y $(= F/DM)$ value. Let y' denote an additional value of Y not included among those used to construct the regression $\hat{Y}$, and let X' be the value of X to which y' corresponds. If y' be an *independent* observation, then

$$(y' - \bar{Y}_{X'}) \quad \text{and} \quad (\hat{Y}' - \bar{Y}_{X'}),$$

where $\hat{Y}'$ denotes the value of $\hat{Y}$ corresponding to $X = X'$, are normally and independently distributed about zero with variances $\sigma^2_{Y\cdot X}$ and $\sigma^2_{\hat{Y}'}$ respectively.

Since the difference of two quantities normally and independently distributed about zero is also distributed normally about zero with variance equal to the sum of the respective variances, it follows that $(y' - \bar{Y}_{X'}) - (\hat{Y}' - \bar{Y}_{X'}) = (y' - \hat{Y}')$ is normally distributed about zero with the variance $\sigma^2_{y\cdot x} + \sigma^2_{\hat{Y}'}$. Using $s^2_{y\cdot x}$ to estimate $\sigma^2_{y\cdot x}$, which is involved in both of these terms, it follows from (C) that

$$t = \frac{y' - \hat{Y}'}{\sqrt{S^2_{\hat{Y}'} + s^2_{y\cdot x}}}, \tag{19}$$

where $\hat{Y}'$ is the value of (4') for $X = X'$ and y' is an additional value of Y for $X = X'$ and $S_{\hat{Y}'}$ the value of $S_{\hat{Y}}$ for $X = X'$, is distributed as Student's t for $N - 2$ degrees of freedom. It should be noticed that here the estimate $s^2_{y\cdot x}$ obtained in connection with $\hat{Y}$ carries all of the burden of estimating $\sigma^2_{Y\cdot X}$. Accordingly, unless our *combined experience with regard to* y' *and* $\hat{Y}'$ is such as would occur 1 time in 20, i.e. unless t of (19) numerically exceeds $t_{.05}$ for $N - 2$ degrees of freedom, it follows that

$$-t_{.05}\sqrt{S^2_{\hat{Y}'} + s^2_{y\cdot x}} \leq y' - \hat{Y}' \leq t_{.05}\sqrt{S^2_{\hat{Y}'} + s^2_{y\cdot x}} \tag{20}$$

which may also be written as

$$\hat{Y}' - t_{.05}\sqrt{S^2_{\hat{Y}'} + s^2_{y\cdot x}} \leq y' \leq \hat{Y}' + t_{.05}\sqrt{S^2_{\hat{Y}'} + s^2_{y\cdot x}}. \tag{21}$$

If, therefore, y' denotes a *future* observation, unless our experience to date (contained in $\hat{Y}$ and $S_{\hat{Y}}$) *and* our future experience with regard to y' are such as to make t of (19) exceed $t_{.05}$ numerically—it being supposed we are sampling a statistically stable universe—then if we predict limits for y' by means of (21) we can associate a confidence of .95 with this *combined* procedure—that is, if we make a habit of evaluating regression lines $\hat{Y}$ and of predicting new observations with their aid by means of (21), then in 95% of the cases in which we take *independent paired steps* of this sort we may expect to be correct with regard to our prediction of y'. It should be noted that if $\hat{Y}$ is "away out" in the first place, which may occur by chance, the combined experience of y' and $\hat{Y}'$ will probably be "away out" too, although y' may be near $\bar{Y}_{X'}$ where it belongs. The 95% wager applies to the combined steps of getting $\hat{Y}$ and y' and *not* to the single step laying off an interval about $\hat{Y}$ in hopes of "catching" y'. In consequence one should not keep on using one regression $\hat{Y}$ over and over again, but should be continually amending "experience to date" as data accumulate.[9]

It should be noted that the above procedure does not yield us an interval which may be expected to include 95% of the future values of y. Such a range

[9] H. Working and H. Hotelling discussed this use of regression to forecast future values, but did not, as far as I can see, emphasize the confidence interval nature of the argument, nor the fact that the probability concerned refers to the two steps involved, and not merely to the latter. The same may be said with regard to Schultz's paper [8].

would be an estimate of the range within which 95% of the population values lie. The difficulties attending the estimation of this type of range are discussed by Shewhart [10], and it appears from his work that in the present state of our knowledge very large samples are required for this purpose. In addition, by a beautiful example, Shewhart shows how a failure to distinguish between confidence intervals associated with a given confidence coefficient, say .95, and intervals containing 95% of the population values, can lead to statements which are quite false.

Recalling to mind that we have been going through all of this reasoning with the aim of finding a way of deciding to what value of C/F $(= X)$ we should tell the dairyman to standardize his milk if he wishes to produce cheese for which F/DM $(= Y)$ is 45 *at least*, we see that our problem consists in getting a lower limit to y' where X' is the value at which we shall advise him to standardize. If, therefore, we leave the right side of the inequality (21) open so that we have

$$\hat{Y}' - t'_{.05}\sqrt{S^2_{\hat{Y}'} + s^2_{y\cdot x}} \leq y', \tag{22}$$

where $t'_{.05}$ is the value of t for which $P\{t < -t'_{.05}\} = .05$, the *sign* of the t value in (19) being considered now, then we seek that value of X, which makes the left side of this equal to 45. For, if y' correspond to this value of X, call it X', then *unless our experience to date plus our future experience* with y' is such as we may expect to occur 1 time in 20 in the long run, y' will be greater than 45, as desired. In other words, we want to solve

$$a + b(X' - \bar{x}) - t'_{.05}\sqrt{s^2_{y\cdot x}\left\{1 + \frac{1}{N} + \frac{(X' - \bar{x})^2}{\Sigma(X - \bar{x})^2}\right\}} = Q \tag{23}$$

for X', where $Q = 45$ in this case. By straightforward algebra the general solution is found to be

$$X' = \bar{x} + \frac{b(Q - a)}{C} \pm \frac{(t'_{.05})s_{y\cdot x}}{C}\sqrt{B(Q - a)^2 + \left(\frac{N + 1}{N}\right)C} \tag{24}$$

in which $a = \bar{y}$, $B = 1/\Sigma(x - \bar{x})^2$, and $C = b^2 - (t^2_{.05})(s^2_{y\cdot x})(B)$, and the sign before the last term is $+$ if b is positive and $-$ if b is negative.

From the data involved in the present problem $N = 149$, $\bar{x} = .8439$

$a = 43.63$, $b = -24.58$, $B = 3.6391$, $s^2_{y\cdot x} = .9448$, $s_{y\cdot x} = .9720$

and for $t'_{.05} = 1.656$, the one-sided 5% value for 147 degrees of freedom, $C = 594.7479$.

Substituting these values in (24) *we find* $X' = .7207$, *and this is the value to which the dairyman should standardize his* C/F *ratio.* If he does, then *unless* the experience to date, leading to $\hat{Y}$ of (11), and the future experience with regard to any new y $(= F/DM)$ value—unless these *combined experiences* are such as to shove the t of (19) beyond the *one-sided* 5% value of t for 147 degrees of freedom *and* in the negative direction, the predicted value of y $(= F/DM)$ will be 45 *at least.* In this sense we may have 95% confidence that our prediction will be correct.

It is clear that the preceding solution can be set up for any desired degree of confidence, say $1 - \epsilon$, by choosing t'_ϵ which is the value of t for which $P\{t < -t'_\epsilon\} = \epsilon$ for the degrees of freedom involved. Furthermore, if an upper limit, instead of a lower limit, were desired, the solution would be the same except for an interchanged usage of the + and − signs before the last term of (24)—for an upper limit one would take a − if b were positive and a + if b were negative. For values of Q not too different from $\bar{y}$ it will usually be possible to find the solution corresponding to the level of confidence desired. However, it is quite possible that a solution may not exist for the value of Q desired, if this be too distant from $\bar{y}$. This difficulty will arise whenever $[(N + 1)/N](t'_\epsilon)^2 s^2_{y \cdot x} B$ is larger than $B(Q - \bar{y})^2 + [(N + 1)/N]b^2$, in which case the radical is imaginary, and no real solution of (24) exists. By graphing the left side of (22) for several values of X' the reason why such cases occur can be readily appreciated.

Since the confidence coefficient in reality relates to the difference $(y' - \hat{Y})$ in which both y' and $\hat{Y}$ are random variables, when applying this method to a particular industrial (or other) problem, one should make repeated $\hat{Y}$ estimates of $\bar{Y}_X$ from time to time in order to insure that the $\hat{Y}$ used is not away off from $\bar{Y}_X$. As mentioned earlier $\hat{Y}$ will assess $\bar{Y}_X$ more accurately if the X values used are spread over a rather wide range—this follows from the nature of (12). By frequent determinations of $\hat{Y}$ even better estimates of $\bar{Y}_X$ can be obtained by pooling the data to date, provided no departures from statistical stability are detected. In this way an increasingly reliable estimate of X' can be determined. By standardizing with $X = X'$ and keeping an eye on the resulting y values, one will be able to see whether this choice of X' is operating satisfactorily. Also, and more important probably, by standardizing $X = X'$ and applying control charts as described by Shewhart [9] and Pearson [5] to the observed y values, one may detect the first signs of a change in conditions "some time before this could be discovered by cruder methods, such as mere inspection of columns of figures."

7. Assaying an Unknown with the Aid of a Previously Established Relationship.

Having come this far, only one step farther is required to obtain a solution to a class of problems having the general nature of the following: A previously calculated regression, $\hat{Y}$, being available, a new value y' is observed and the value of X, say X', to which it corresponds has been lost sight of, or was never known. What value of X should be taken as the best single estimate of X', and within what limits can we assess X' with a confidence coefficient of .95 say?

From our previous discussion it is clear, I think, that in repeated sampling of *both* $\hat{Y}$ and y' the inequality (21) should hold 95% of the time, if $t_{.05}$ is the value for which $P\{|\, t \,| > t_{.05}\} = .05$. Accordingly, *unless* our present experience with regard to $\hat{Y}$ *and* y' is in the upper or in the lower .025 tail of the t-distribution, y' is related to $\hat{Y}$ as indicated by (21). But the left side of (21) is really

the same as the left side of (23) with $t_{.05}$ in place of $t'_{.05}$, and the right side of (21) can likewise be obtained from the left side of (23) by replacing $t'_{.05}$ by $-t_{.05}$, and in both cases y' corresponds to Q, X' being unknown as in the previous problem. In short, by setting $Q = y'$ in (24) and replacing $t'_{.05}$ by $t_{.05}$, we can use this revised (24) to obtain upper and lower limits for X', and *unless* our combined experience with regard to $\hat{Y}$ and y' is such as would occur 1 time in 20, the value of X which truly corresponds to y' will be within these limits. The "best" single estimate will be $X' = \bar{x} + \dfrac{y' - \bar{y}}{b}$, which can be obtained from (24) by setting $t = 0$, and it should be noted that the upper and lower limits of X' for a given confidence level are *not* symmetrical with respect to this value. With regard to the data of Figure 2, if our new value $y' = 45$, and if the confidence desired were merely .90 (so that we can use $t'_{.05} = t_{.10}$), the calculations yield $.7207 \leq X' \leq .8539$ with $X' = .7884$ as the best single estimate.

It is unlikely that a dairyman would ever be interested in obtaining limits for C/F from the F/DM value of a finished cheese, so that he would probably never have any use for this additional technique. On the other hand the preceding situation is a common one in connection with problems of biological assay where it is desired to evaluate the potency of a substance by comparing the response it produces, when administered to one or more animals, with a dosage-response relation previously established with dosages of known strength. In the preceding problem we considered the case in which y' was a single additional observation corresponding to an unknown X'. If, instead, we had $\bar{y}'$, the mean value of N' additional observations corresponding to an unknown X', it is clear that the denominator of (19) will be $\sqrt{S^2_{\hat{Y}'} + s^2_{y\cdot x}/N'}$ in this case, so that confidence limits for X' corresponding to a confidence coefficient of .95 will be

$$X' = \bar{x} + \frac{b(\bar{y}' - \bar{y})}{C} \pm \frac{t_{.05}\cdot s_{y\cdot x}}{C}\sqrt{B(\bar{y}' - \bar{y})^2 + \left(\frac{N + N'}{NN'}\right)C} \tag{25}$$

and the "best" single estimate of X' will be

$$X' = \bar{x} + \frac{\bar{y}' - \bar{y}}{b}, \tag{26}$$

where $\bar{y}$ is the mean of the y's in the analysis of the original N values;
$\bar{y}'$ " " " " the additional N' y's corresponding to the unknown X';
b " " regression coefficient in (4′);
$s_{y\cdot x}$ is given by (14) and depends on the scatter of the original y values about the regression Y, and is based on $N - 2$ degrees of freedom;
$B = 1/\Sigma(x - \bar{x})^2$, the summation being over the original X values;
$t_{.05}$ = two-sided 5% significance level of t for $N - 2$ degrees of freedom;
and $C = b^2 - t^2_{.05}\cdot s^2_{y\cdot x}\cdot B$.

In practice N' is usually small compared with N, so that $s^2_{y\cdot x}$ based on the original analysis will probably be used. However, if it is desired to make use of the dispersion of the new y' values to "improve" the estimate of $\sigma^2_{Y\cdot X}$, then $\bar{S}^2_{y\cdot x} = [(N-2)s^2_{y\cdot x} + (N'-1)s'^2]/(N+N'-3)$ should be used in place of $s^2_{y\cdot x}$, where $s'^2 = \Sigma(y' - \bar{y}')^2/(N'-1)$, and the $t_{.05}$ value corresponding to $(N+N'-3)$ degrees of freedom used. Mathematically this is preferable to the above, but involves considerably more calculating, and probably would not be used by the practical man.

We shall illustrate the use of (25) and (26) in connection with the data of Figure 3 obtained from autopsies of 69 rats which had received doses of estradiol varying from 0.025 micrograms to .2 micrograms.[10] It was found that a linear relation, with a common variance on the various dosages, existed between $X = \log_{10}$ dose and $Y = \sqrt{\text{uterine wt.}}$ These are the quantities portrayed in Figure 3. The least squares line is

$$\hat{Y} = 6.9023 + 3.4004(X + 1.0777) = 10.567 + 3.400X, \tag{27}$$

and is seen to be a good fit.

Carrying through the necessary calculations we find that 95% confidence limits for X', the true log dose, corresponding to a mean response of $\bar{y}'$ based on N' values, are

$$X' = -1.0777 + 0.2964(\bar{y}' - 6.9023) \pm .07074\sqrt{.1376(\bar{y}' - 6.9023)^2 + \left(\frac{69+N'}{69\times N'}\right)(.09062)} \tag{28}$$

and the optimum single estimate is

$$X' = 0.2941\,\bar{y}' - 3.1077. \tag{29}$$

Dr. C. I. Bliss has informed me in correspondence that seldom is the sensitivity of an animal species to a hormone or other drug constant enough for the actual procedure outlined above to be reliable, so that in assaying any given sample it should always be tested in parallel with a standard preparation. If the slope of the regression, i.e. b, is fairly stable, even though the position changes, it is possible to assay the relative strength of an unknown by administering it and a standard at a single dilution, but it is preferable to use at least two dilutions in each assay so that it may be discovered whether the new b agrees substantially with that given by the standard dosage-response curve. Discussions of the procedures to be used in these cases will be found in references [26] to [31] from which I have received much inspiration.

[10] These data have been discussed by Lauson, Heller, Golden, and Sevringhaus [32] of the Wisconsin General Hospital, to whom I extend thanks for permission to use them in the present paper. Only a portion of their data have been used as the linear relation discussed below failed outside of the dosage limits given above.

8. **Concluding Remarks.** The formulae and ideas presented in this paper have been drawn in the main from the articles and books listed at the end of this paper. By turning to these references the reader often will find a fuller account of methods and applications than has been given here. In many cases

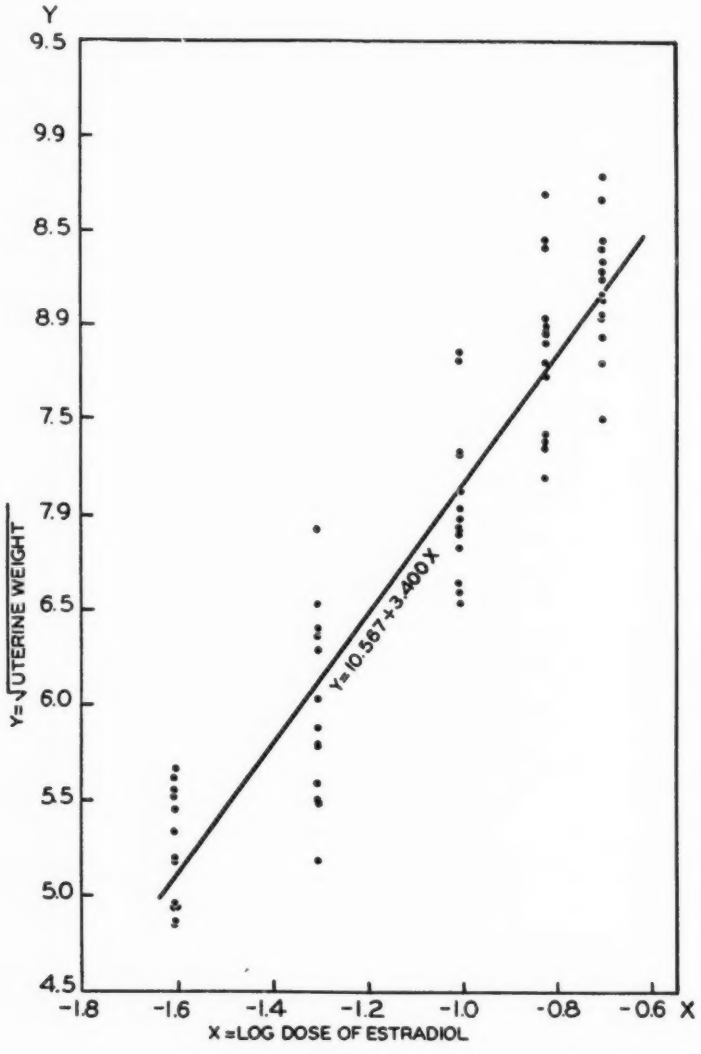

FIG. 3

the reader will find that the author of one of the references has placed emphasis on getting the answer. In the present paper the emphasis has been on the ideas and assumptions involved, the aim being to promote understanding of the methods discussed. In particular, the following two points have been stressed here:

(a) When the values of one of a pair of random variables are selected by the research worker, or when one of the variables is allowed to take values in only a restricted portion of its real range, then inferences with regard to an unknown value of this variable, say X, based on the corresponding (known) value of the other variable, say Y, are mathematically valid only when inferred from the relationship giving Y as a function of X; and

(b) The resulting inference is in the form of a confidence interval whose confidence coefficient is associated with the joint experience consisting of the observed regression of Y on X *and* the observed (or future) additional sample involving the unknown value of X, and *not* merely with the latter.

The ideas and assumptions which have been discussed have been illustrated on two examples.

Closer coöperation is possible between the practical man and the statistical theorist when the latter fully appreciates the problems of the former, *and* when the former, in turn, understands the methods advocated by the latter.

THE UNIVERSITY OF WISCONSIN.

REFERENCES

General Literature

[1] W. EDWARDS DEMING, *Some Notes on Least Squares.* Washington: Graduate School, Dept. of Agric., 1938.

[2] W. EDWARDS DEMING, "Some Thoughts on Curve Fitting and the Chi-Test," *Jour. American Statistical Association,* vol. 33 (1938), pp. 543–551.

[3] R. A. FISHER, *Statistical Methods for Research Workers.* Edinburgh: Oliver & Boyd, 6th edition, 1936.

[4] R. A. FISHER, "Applications of 'Student's' Distribution," *Metron,* vol. 5 (1925), pp. 90–104.

[5] E. S. PEARSON, *The Application of Statistical Methods to Industrial Standardization and Quality Control.* London: British Standards Institution, 1935.

[6] P. R. RIDER, "A Survey of the Theory of Small Samples," *Annals of Mathematics,* vol. 3 (1930), pp. 577–628. Copies available from Nordemann Publishing Co., N. Y. C.

[7] P. R. RIDER, *An Introduction to Modern Statistical Methods.* New York: John Wiley & Sons, 1939. pp. 78, 79, 90, 91.

[8] HENRY SCHULTZ, "The Standard Error of a Forecast from a Curve," *Jour. American Statistical Association,* vol. 25 (1930), pp. 139–185.

[9] W. A. SHEWHART, *Economic Control of Quality of Manufactured Product.* New York: D. Van Nostrand, 1931.

[10] W. A. SHEWHART, *Statistical Methods from the Viewpoint of Quality Control.* Washington: Graduate School, U. S. D. A. In press.

[11] "Student" (the late W. S. GOSSET), "The Probable Error of a Mean," *Biometrika,* vol. 6 (1908), pp. 1–25.

[12] HOLBROOK WORKING AND HAROLD HOTELLING, "Applications of the Theory of Error to the Interpretation of Trends," *Jour. American Statistical Association,* vol. 24 (1929), pp. 73–85.

Literature Relating to Fiducial Limits and Confidence Intervals

[13] R. A. FISHER, "Inverse Probability," *Proc. Cambridge Philosophical Society,* vol. 26 (1930), pp. 528–535.

[14] R. A. FISHER, "The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters," *Proc. Royal Society of London*, Series A, vol. 139, pp. 343–348.

[15] R. A. FISHER, "The Fiducial Argument in Statistical Inference," *Annals of Eugenics*, vol. 6 (1935), pp. 391–398.

[16] J. NEYMAN, "On the Two Different Aspects of the Representative Method," *Jour. Royal Statistical Society*, vol. 97 (1934), pp. 558–625.

[17] J. NEYMAN, "On the Problem of Confidence Intervals," *Annals of Mathematical Statistics*, vol. 6 (1935), pp. 111–116.

[18] J. NEYMAN, "Outline of a Theory of Estimation Based on the Classical Theory of Probability," *Phil. Trans. Roy. Soc. London*, Series A, vol. 236 (1937), pp. 333–380.

[19] J. NEYMAN, *Lectures and Conferences on Mathematical Statistics*; edited by Deming; Washington: Graduate School, Dept. of Agric., 1938.

[20] J. NEYMAN, *L'estimation statistique traitée comme un problème classique de probabilité.* Actualités scientifiques et industrielles, Nr. 739 (1938), pp. 25–57. Paris: Hermann et Cie.

[21] H. L. RIETZ, "On a Recent Advance in Statistical Inference," *Amer. Math. Monthly*, vol. 45, pp. 149–158.

[22] B. L. WELCH, "On Confidence Limits and Sufficiency with Particular Reference to Parameters of Location," *Annals of Mathematical Statistics*, vol. 10 (1939), pp. 58–69.

[23] ALBERT WERTHEIMER, "A Note on Confidence Intervals and Inverse Probability," *Annals of Mathematical Statistics*, vol. 10 (1939), pp. 74–76.

[24] S. S. WILKS, "Fiducial Distributions in Fiducial Inference," *Annals of Mathematical Statistics*, vol. 9 (1938), pp. 272–280.

[25] E. B. WILSON, "Probable Inference, the Law of Succession, and Statistical Inference," *Jour. American Statistical Association*, vol. 22 (1937), pp. 209–212.

Literature Cited Relating to Methods of Biological Assay

[26] C. I. BLISS, "The Calculation of the Dosage-Mortality Curve," *Annals of Applied Biology*, vol. 22 (1935), pp. 134–167.

[27] C. I. BLISS, "The Comparison of Dosage-Mortality Data," *Annals of Applied Biology*, vol. 22 (1935), pp. 307–333.

[28] C. I. BLISS, "The Determination of the Dosage-Mortality Curve from Small Numbers," *Quarterly Journal of Pharmacy and Pharmacology*, vol. 11 (1938), pp. 192–216.

[29] J. H. BURN, *Biological Standardization.* Oxford Univ. Press, 1937.

[30] J. H. GADDUM, (British) Medical Research Council Special Report Ser. No. 183 (1933).

[31] J. O. IRWIN, "Statistical Methods Applied to Biological Assays," *Supplement Jour. Royal Statistical Society*, vol. 4 (1937), pp. 1–48.

[32] H. D. LAUSON, CARL G. HELLER, JUNE B. GOLDEN AND E. L. SEVRINGHAUS, "The Immature Rat Uterus in the Assay of Estrogenic Substances, and a Comparison of Estradiol, Estrone and Estriol," *Endocrinology*, vol. 24 (1939), pp. 35–44.

NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

NOTE ON THE L_1 TEST FOR MANY SAMPLES

By A. M. Mood

Neyman and Pearson[1] have discussed a method for testing the hypothesis that k samples have been drawn from normal populations with the same variances by means of a statistical function, L_1, defined by

$$L_1^{\frac{2}{N}} = \prod_{t=1}^{k} \left(\frac{s_t^2}{s^2}\right)^{\frac{n_t}{2}}$$

where n_t is the number of elements in the t-th sample, s_t^2 is the sample variance and

$$s^2 = \sum_{t=1}^{k} \frac{n_t}{N} s_t^2 \qquad N = \sum_{t=1}^{k} n_t .$$

For convenience, we shall denote $L_1^{\frac{2}{N}}$ by λ. In their paper Neyman and Pearson have found the moments of λ and have shown that the distribution of $-2 \log_e \lambda$ approaches that of χ^2 with $k - 1$ degrees of freedom when the number of elements in each of the k samples becomes large. In some applications of this test the question arises as to whether the χ^2 law is a good approximation when the number of samples is large in comparison with the number of elements in each sample. For example, in a certain educational study, the number of schools was much greater than the number of pupils in each school, and it was desired to test for heterogeneity of variances of scores on a given examination using L_1 as the criterion. The purpose of this note is to examine the behavior of the L_1 test for large values of k.

Wilks has obtained the distribution of λ as a definite integral; it is, however, a rather cumbersome form to handle. The procedure here will be simply to compare the first few semi-invariants of $-2 \log \lambda$ with those of χ^2. The p-th moment of λ is[2]

$$\mu'(p) = \frac{N^{\frac{pN}{2}} \Gamma\left(\frac{N-k}{2}\right)}{\Gamma\left(\frac{(p+1)N-k)}{2}\right)} \prod_{t=1}^{k} \frac{\Gamma\left(\frac{(p+1)n_t - 1}{2}\right)}{n_t^{\frac{p n_t}{2}} \Gamma\left(\frac{n_t - 1}{2}\right)} . \tag{1}$$

[1] "On the problem of k Samples," *Bulletin de l'Académie Polonaise des Sciences et des Lettres*, Série A (1931), pp. 460–481.

[2] Ibid., p. 472.

Since

$$E(e^{(-2\log\lambda)\theta}) = E(\lambda^{-2\theta})$$

the characteristic function of $-2\log\lambda$ is obtained on replacing p by -2θ in (1), where $\theta = it$, t being a real variable. The logarithm of the characteristic function is the generating function of the semi-invariants; denoting the latter by $\psi(\theta)$, we have

$$\psi(\theta) = \log \mu'(-2\theta). \tag{2}$$

After substitution of (1) in (2), the resulting expression can be simplified by means of the Weierstrass factored form of $1/\Gamma(x)$ which is

$$\frac{1}{\Gamma(x)} = xe^{\gamma x}\prod_{r=1}^{\infty}\left(1+\frac{x}{r}\right)e^{-\frac{x}{r}},$$

where γ is the Euler constant .577. The final result is

$$\psi(\theta) = \theta\left[\sum_{t=1}^{k} n_t \log n_t - N\log N\right] + \sum_{r=0}^{\infty}\log\frac{2r+N'-k}{2r+N-k} - \sum_{t=1}^{k}\sum_{r=0}^{\infty}\log\frac{2r+n_t'-1}{2r+n_t-1} \tag{3}$$

where $N' = N(1-2\theta)$ and $n_t' = n_t(1-2\theta)$.

The semi-invariants of $-2\log\lambda$ are given by the derivatives of $\psi(\theta)$ evaluated at $\theta = 0$; these will be denoted by $\lambda_1, \lambda_2, \cdots$. λ_1 and λ_2 are the mean and variance respectively, and in general the semi-invariants are related to the moments, μ_s', by[3]

$$\mu_s' = \sum_{i=1}^{s}\binom{s-1}{i-1}\lambda_i\mu_{s-i}'. \tag{4}$$

From the generating function (3) we obtain:

$$\lambda_1 = \psi'(0) = \sum_{t=1}^{k} n_t\log n_t - N\log N - \sum_{r=0}^{\infty}\frac{2N}{2r+N-k} + \sum_{t=1}^{k}\sum_{r=0}^{\infty}\frac{2n_t}{2r+n_t-1} \tag{5}$$

$$\lambda_s = \psi^{(s)}(0) = (s-1)!\sum_{r=0}^{\infty}\left[\sum_{t=1}^{k}\frac{(2n_t)^s}{(2r+n_t-1)^s} - \frac{(2N)^s}{(2r+N-k)^s}\right] \qquad s = 2, 3, \cdots \tag{6}$$

[3] See e.g., Charles Jordan, *Statistique Mathématique*, p. 41.

The infinite sums can be well approximated by integration when the n_i are moderately large, giving

$$\lambda_1 = \sum_{i=1}^{k} n_i \log \frac{n_i(N-k-1)}{N(n_i-2)} \tag{7}$$

$$\lambda_s = (s-2)!2^{s-1}\left[\sum_{i=1}^{k} \frac{n_i^s}{(n_i-2)^{s-1}} - \frac{N^s}{(N-k-1)^{s-1}}\right] \quad s = 2, 3, \cdots \tag{8}$$

and when the samples are of equal size, that is

$$n_1 = n_2 = \cdots = n_k = n, \qquad N = kn$$

equations (7) and (8) become

$$\lambda_1 = kn \log\left(1 + \frac{k-1}{k(n-2)}\right) \tag{9}$$

$$\lambda_s = (s-2)!2^{s-1}\left[\frac{kn^s}{(n-2)^{s-1}} - \frac{k^s n^s}{(kn-k-1)^{s-1}}\right] \quad s = 2, 3, \cdots \tag{10}$$

It is worth noting that these last two expressions are monotonic decreasing functions of n for a fixed $k > 1$; hence when the sample sizes are unequal the true values of the λ_s lie between the values given by substituting the least and greatest n_i for n in (9) and (10). This fact supports the suggestion of Nayer[4] on page 47 of his paper on the application of the L_1 test. He has computed tables for the critical values of L_1 when the sample sizes are equal, and suggests that when the sizes are unequal but not radically different, the average value of n_i may be used.

The limiting values given by

$$\lambda_s \xrightarrow[n \to \infty]{} (s-1)!2^{s-1}(k-1) \qquad s = 1, 2, 3, \cdots \tag{11}$$

are the semi-invariants of χ^2 with $k-1$ degrees of freedom as is easily verified by induction using (4) and the following expression for the moments of χ^2 with m degrees of freedom:

$$\mu_s' = m(m+2)(m+4)\cdots(m+2s-2).$$

For a fixed $n > 2$ the quantities

$$\frac{\lambda_s}{(s-1)!2^{s-1}(k-1)}$$

are monotonic decreasing functions of k, however the variation is rather slight as is evident from the following table:

[4] "An investigation into the Application of the Neyman and Pearson L_1 Test, with Tables of Percentage Limits," *Statistical Research Memoirs*, Vol. I (1936), pp. 38-51.

n	20		100		∞	
k	10	∞	10	∞	10	∞
$\frac{\lambda_1}{k-1}$	1.084	1.081	1.016	1.015	1	1
$\frac{\lambda_2}{2(k-1)}$	1.176	1.170	1.032	1.031	1	1
$\frac{\lambda_3}{8(k-1)}$	1.275	1.265	1.048	1.046	1	1
$\frac{\lambda_4}{48(k-1)}$	1.384	1.369	1.065	1.062	1	1

These results indicate that the degree of approximation of $-2 \log \lambda$ to the χ^2 law with $k-1$ degrees of freedom is mainly dependent on n, and is for all practical purposes independent of k when n is moderately large.

PRINCETON UNIVERSITY,
PRINCETON, NEW JERSEY.

ON TCHEBYCHEFF APPROXIMATION FOR DECREASING FUNCTIONS

BY C. D. SMITH

The problem of estimating the value of a probability by means of moments of a distribution function has been studied by Tchebycheff, Pearson, Camp, Meidel, Narumi, Markoff, and others. Approximations without regard to the nature of the function have not been very close. However the closeness of the approximation has been materially improved by placing certain restrictions on the nature of the distribution function.[1] For example, when $y = f(x)$ is an increasing function from $x = 0$ to $x = c\sigma$ and a decreasing function beyond that point, the corresponding probability function $y = P_x$ is concave downward from $x = 0$ to $x = c\sigma$ and concave upward beyond that point. Here P_x is the probability that a variate taken at random from the distribution will fall at a distance at least as great as x from the origin. Beginning with these conditions I have established the inequality[1]

[1] B. H. Camp, "A New Generalization of Tchebycheff's Statistical Inequality", *Bulletin of the American Mathematical Society*, Vol. 28, (1922), pp. 427-32.

C. D. Smith, "On Generalized Tchebycheff Inequalities in Mathematical Statistics," *The American Journal of Mathematics*, Vol. 52, (1930), pp. 109-26.

$$P_x \leqq \frac{\beta_{2r-2} - c^{2r}[(2rP_{c\sigma} + 1)/(2r + 1)]}{(t/\theta)^{2r} - c^{2r}}; \qquad x = t\sigma, \qquad \beta_{2r-2} = \frac{M_{2r}}{\sigma^{2r}}.$$

(1)

$$t = \frac{2r}{2r + 1} \cdot \frac{t^{2r+1} - (c\theta)^{2r+1}}{\theta[t^{2r} - (c\theta)^{2r}]}.$$

The upper bound was obtained by substituting $P_{t\sigma}$ for $P_{c\sigma}$, $t > c$, and the special values $c = r = 1$, and $t = 2$, gave the result $P_{2\sigma} \leqq .092$.

The purpose of this paper is to give an estimate of $P_{c\sigma}$ which will substantially improve the approximation to the value of P_x obtained from (1). Let $y = f(x)$ be a monotonic increasing function from $x = 0$ to $x = c\sigma$ and a monotonic decreasing function from $x = c\sigma$ to the upper end of the range of x. With P_x as the probability that a variate taken at random from the distribution will deviate from the origin by an amount at least x we know that the curve of $y = P_x$ is concave downward from $x = 0$ to $x = c\sigma$ and concave upward beyond that point. When $y = f(x)$ is of finite range the probability curve and the curve of $y = f(x)$ will terminate at the same point on the x-axis. The probability curve will approach the x-axis when the range of the function is infinite. In either case we may take a distribution $y = g(x)$ to follow the curve of the given function from $x = 0$ to $x = c\sigma$ and to follow a horizontal line from $x = c\sigma$ to a finite distance A from the orign and such that the area under the curve is the same as that under the curve of $y = f(x)$. Obviously the probability curve for $y = g(x)$ will be a straight line from $(c\sigma, P_{c\sigma})$ to the point $(A, 0)$, and since the curve of $y = P_x$ is concave upward beyond $(c\sigma, P_{c\sigma})$ it will remain below a straight line to a point very near $(A, 0)$. Also it is evident that the straight line has a y-intercept greater than unity since a straight line beginning at $(0, y)$ and extending a distance A from the origin would give a probability function whose graph follows the straight line from $(0, 1)$ to $(A, 0)$. Obviously the ordinates of this probability graph for values of x in the interval from $x = 0$ to $x = c\sigma$ are less than the corresponding ordinates for the curve which increases for x in the same interval and then follows the horizontal line. Hence a line through points $(0, 1)$ and $(c\sigma, P_{c\sigma})$ is above the line $(c\sigma, P_{c\sigma})$ to $(A, 0)$ for all points beyond $c\sigma$.

We may use the line through points $(0, 1)$ and $(c\sigma, P_{c\sigma})$ as a basis for estimating $P_{c\sigma}$ in (1). The equation of the line is $y = \dfrac{P_{c\sigma} - 1}{c\sigma} x + 1$ with x-intercept $\dfrac{c\sigma}{1 - P_{c\sigma}}$. The line remains above the curve of $y = P_x$ from $x = c\sigma$ to a point very near the x-intercept and so we may use the line from $x = c\sigma$ to the crossing point. The range of validity seems to be sufficient for practical use since $P_{c\sigma}$ is usually near .9 and c is a fraction. For $P_{c\sigma} = .9$, $c = .5$, the intercept of the line is approximately 5σ. Let the ordinate under the line be y_t, $(t > c)$, and then $P_{c\sigma} = 1 + \dfrac{c}{t}(y_t - 1)$. For the probability curve $y = P_x$ we have $P_{c\sigma} >$

$1 + \frac{c}{t}(P_{t\sigma} - 1)$ since $y_t > P_{t\sigma}$. Substitution of $1 + \frac{c}{t}(P_{t\sigma} - 1)$ for $P_{c\sigma}$ in (1) gives

$$(2) \qquad P_{t\sigma} \leqq \frac{\beta_{2r-2} - c^{2r}\left(1 - \dfrac{2rc}{t(2r+1)}\right)}{\left(\dfrac{t}{\theta}\right)^{2r} - c^{2r}\left(1 - \dfrac{2rc}{t(2r+1)}\right)}, \qquad t\sigma < \frac{c\sigma}{1 - P_{c\sigma}}, \qquad \theta \text{ as in (1).}$$

To indicate the amount of improvement let $c = r = 1$, and $t = 2$. From (1) $P_{2\sigma} \leqq .092$ while from (2) $P_{2\sigma} \leqq .056$. One may work from any origin other than the mean by letting $h = c\sigma$ in (2).

MISSISSIPPI STATE COLLEGE.

CORRECTION OF SAMPLE MOMENT BIAS DUE TO LACK OF HIGH CONTACT AND TO HISTOGRAM GROUPING

BY DINSMORE ALTER

The first correction of sample moment bias was devised by W. F. Sheppard [1]. His method corrects for histogram grouping on the assumption of high contact at both ends of the frequency curve. Usually this is a sufficient correction. In some cases, however, of J-shaped curves the error remaining is even more serious than in the original histogram moments.

A method developed by E. Pairman and Karl Pearson [2] makes a complete correction for both of these sources of bias. The only advantage claimed for the method to be developed here over theirs lies in simplicity of mathematical theory.

A third correction is given by Elderton [3]. In his method he assumes that there is no error due to histogram grouping and he develops a correction for lack of high contact, in so far as the zero-th moment is concerned. The following work may be thought of largely as an extension of his method although it will have certain variations.

Let A_x and ν'_m be defined as follows,

$$A_x \equiv \int_{t=-\frac{1}{2}}^{+\frac{1}{2}} y_{x+t}\, dt$$

$$\nu'_m \Sigma A_x \equiv \Sigma x^m A_x$$

The definite integrals are the areas of the histogram rectangles if a scale of x be chosen to reduce their width to unity. Let μ'_m be defined by

$$\mu'_m \Sigma A_x \equiv \int_{l_1}^{l_2} x^m y_x\, dx$$

In the first equation the x's form a series of equally spaced constants. In the second, x is a continuous variable. The summations are to extend over the equally spaced values of x.

If the data form a histogram, l_1 and l_2 are respectively the values of x at the left edge of the left-hand rectangle and the right edge of the right-hand one. If the data are the values of y_x at isolated points, l_1 is the value of x one-half unit smaller than the smallest value given in the sample and l_2 is one-half greater than the largest. It would be perfectly satisfactory, of course, to define these limits differently. As defined, however, they parallel the histogram case. Distributions of this latter type will be called *point frequency distributions.*

As is customary, the primed moments denote those about an arbitrary origin. Moments corrected for lack of high contact and for grouping will be denoted by μ'_m or by μ_m if taken about the mean. Numerical raw moments will be denoted by ${}_n\mu'_m$. There are two entirely different methods of approach to this bias problem.

(a) The bias may be put into the algebraic form of the frequency curve and equated directly to the numerical raw moments. In the case of a point frequency distribution such a method forms the algebraic values of y_x for each point given in the sample and, therefore, puts the raw moments into algebraic form to be equated to the numerical ones. This is the simplest method of correction if the distribution is a power series. For most types the method leads into difficulties which complicate it beyond practical use.

(b) The raw moments given, whether ${}_n\mu'_m$'s or ν'_m's can be corrected to approximate very closely the desired μ'_m's as defined above.

A point frequency distribution gives ${}_n\mu'_m \equiv \Sigma x^m y_x$. If there is high contact ${}_n\mu'_m$ is an unbiased observed estimate of μ'_m. This second form of method will be developed here primarily as a correction to ${}_n\mu'_m$.

Only one assumption is involved. Fifth differences of y_x will be considered as negligible. Any interpolation formula is available but Stirling's will be employed.

$$y_{x+t} = y_x + t\Delta'_x + \frac{t^2}{2}\Delta''_x + \frac{t(t^2-1)}{6}\Delta'''_x + \frac{t^2(t^2-1)}{24}\Delta^{\text{iv}}_x$$

$$\Delta'_x \equiv \tfrac{1}{2}(\Delta'_{x-\frac{1}{2}} + \Delta'_{x+\frac{1}{2}}), \qquad \Delta'''_x \equiv \tfrac{1}{2}(\Delta'''_{x-\frac{1}{2}} + \Delta'''_{x+\frac{1}{2}})$$

$$\mu'_m \Sigma A_x \equiv \int_{l_1}^{l_2} x^m y_x\,dx = \Sigma \int_{-\frac{1}{2}}^{+\frac{1}{2}} (x+t)^m y_{x+t}\,dt$$

$$= \Sigma \int_{-\frac{1}{2}}^{+\frac{1}{2}} \left[x^m + mx^{m-1}t + \frac{m(m-1)}{2}x^{m-2}t^2 + \cdots + t^m\right] y_{x+t}\,dt$$

$$\mu'_m = \nu'_m + \frac{1}{\Sigma A_x}\Sigma\left[mx^{m-1}\int_{-\frac{1}{2}}^{+\frac{1}{2}} t y_{x+t}\,dt + \frac{m(m-1)}{2}x^{m-2}\int_{-\frac{1}{2}}^{+\frac{1}{2}} t^2 y_{x+t}\,dt + \cdots\right]$$

Using Stirling's formula:

$$\mu'_m = \nu'_m + \frac{1}{\Sigma A_x}\Sigma\left\{y_x\left[\frac{m(m-1)}{24}x^{m-2} + \frac{m(m-1)(m-2)(m-3)}{1920}x^{m-4}\right]\right.$$
$$\left. + \Delta'_x\left[\frac{mx^{m-1}}{12} + \frac{m(m-1)(m-2)x^{m-3}}{480}\right]\right.$$

$$+\Delta_x''\left[\frac{m(m-1)x^{m-2}}{320}+\frac{m(m-1)(m-2)(m-3)x^{m-4}}{21504}\right]$$
$$-\Delta_x'''\left[\frac{17\,mx^{m-1}}{1440}+\frac{23m(m-1)(m-2)x^{m-3}}{80640}\right]$$
$$-\Delta_x^{\text{iv}}\left[\frac{23m(m-1)x^{m-2}}{107520}+\frac{29m(m-1)(m-2)(m-3)x^{m-4}}{9289728}\right]$$
$$+\text{terms involving } (m-4)\Bigg\}$$

From this

$$\mu_1'=\nu_1'+\frac{1}{\Sigma A_x}\Sigma\left[\frac{\Delta_x'}{12}-\frac{17\Delta_x'''}{1440}\right]$$

$$\mu_2'=\nu_2'+\frac{1}{\Sigma A_x}\Sigma\left[\frac{y_x}{12}+\frac{x\Delta_x'}{6}+\frac{\Delta_x''}{160}-\frac{17x\Delta_x'''}{720}-\frac{23\Delta_x^{\text{iv}}}{53760}\right]$$

$$\mu_3'=\nu_3'+\frac{1}{\Sigma A_x}\Sigma\left[\frac{xy}{4}+\left(\frac{1}{80}+\frac{x^2}{4}\right)\Delta_x'+\frac{3x}{160}\Delta_x''-\left(\frac{17x^2}{480}+\frac{23}{13440}\right)\Delta_x'''-\frac{23}{17920}\Delta_x^{\text{iv}}\right]$$

$$\mu_4'=\nu_4'+\frac{1}{\Sigma A_x}\Sigma\left[\left(\frac{x^2}{2}+\frac{1}{80}\right)y_x+\left(\frac{x^3}{3}+\frac{x}{20}\right)\Delta_x'+\left(\frac{3x^2}{80}+\frac{1}{896}\right)\Delta_x''-\left(\frac{17x^3}{360}+\frac{23x}{3360}\right)\Delta_x'''-\left(\frac{23x^2}{8960}+\frac{29}{387072}\right)\Delta_x^{\text{iv}}\right].$$

Ordinarily it will not be necessary to use all of the corrective terms.

For point frequency distributions the application of these equations is direct. The ν_m''s may be computed from

$$A_x=y_x+\frac{\Delta_x''}{24}-\frac{17\Delta_x^{\text{iv}}}{5760},$$

and the definition of ν_m'. There is, however, a theoretical difficulty in a case for which the data have been given as a histogram. In such a case the values of A_x are all that have been known originally. The Δ's are not the ones demanded by the equation. The relationship to the proper ones is simple:

$$\Delta_{A_x}'=\Delta_x'+\frac{\Delta_{x+1}''-\Delta_{x-1}''}{24}-\frac{15}{5760}(\Delta_{x+1}^{\text{iv}}-\Delta_{x-1}^{\text{iv}}),\text{ etc.}$$

It is possible to compute the Δ_x^{i}'s from this equation but the discrepancy is small and moreover the Δ^{i}'s are used only in corrective terms. Probably the error involved by use of the wrong Δ^{i}'s is negligible in any actual case of data that ever will be studied. In the numerical example to follow, the very slight errors remaining in the μ_m''s are due, probably, to this neglect.

Pairman and Pearson gave a numerical example in which both the lack of high contact and the grouping introduced large errors. They started with $y_x = 100{,}000\ \sqrt{x}$ and from this formed ten values of A_x. From these they computed the ν_m''s and corrected them to get the μ_m''s. The exact values of the latter were already known to them through integration of the original equation.

The following table compares four values of moments from these data.

m	ν_m'	μ_m' by Sheppard's Formula	μ_m' with Pairman-Pearson Full Corrections	Method Developed Here	True Values
1	5.9880	5.9880	5.9994	5.9996	6.0000
2	42.6900	42.6067	42.8570	42.8576	42.8571
3	331.0854	329.5884	333.3349	333.3387	333.3333
4	2698.7735	2677.4576	2727.2757	2727.3555	2727.2727

Despite the use of the $\Delta_{A_x}^i$'s instead of Δ_x^i's, the results of this method are almost as good as by the older one. The method has the additional advantage of unifying the theories of the correction of moments from the two types of distribution.

REFERENCES

[1] W. F. Sheppard, "Calculation of the Most Probable Values of Frequency Constants," *Proceedings of the London Mathematical Society*, Vol. XXIX, pp. 353–380, 1898.

[2] E. Pairman and Karl Pearson, "Corrections for Moment Coefficients," *Biometrika*, Vol. XII, page 231 et seq.

[3] W. Palin Elderton, *Frequency Curves and Correlation*, pp. 24–27, Charles and Edwin Layton, London, 2d Edition, 1927.

GRIFFITH OBSERVATORY,
AND
CALIFORNIA INSTITUTE OF TECHNOLOGY.

FREQUENCY DISTRIBUTION OF PRODUCT AND QUOTIENT

BY E. V. HUNTINGTON

The main purpose of this note is to establish Theorems 1 and 2. For the sake of completeness, the more familiar Theorems 3 and 4 are appended. All four of these theorems have numerous applications in the theory of frequency distributions. While the proofs of Theorems 1 and 2 in the elementary forms here given (and used in my class-room notes since 1934) can hardly be new, they seem not to be readily accessible in the current text-books.

THEOREM 1. *Suppose a variable x is distributed in accordance with a probability law* $\int_0^\infty f(x)dx = 1$; *and a variable y in accordance with a probability law* $\int_0^\infty F(y)dy$

= 1, x and y being independently distributed. Then the product, u = xy, will be distributed according to the law $\int_0^\infty P(u)du = 1$, *where*

$$P(u) = \int_0^\infty f(u/y)\, F(y)(1/y)\, dy.$$

(The definite integral is a convenient representation of a probability law, since the limits on the integral sign indicate the interval over which the probability law is defined.)

PROOF. Represent the distribution of x by the density of dots along the axis of x, and the distribution of y by the density of dots along the axis of y. Since, by definition, the (relative) number of dots in an interval dx is $f(x)dx$ and the (relative) number of dots in an interval dy is $F(y)dy$, and since each dot in the interval dx is paired with each dot in the interval dy (in accordance with the hypothesis of independence), it follows that the (relative) number of dots in the corresponding area $dxdy$ will be $[f(x)dx][F(y)dy]$.

Now for fixed values of u and Δu, plot the curves $xy = u$ and $xy = u + \Delta u$ in the xy plane, as shown in Figure 1. Then the (relative) number of dots in the area bounded by these two curves is precisely what is meant by $P(u)\Delta u$. Hence the expression $P(u)\Delta u$ may be built up by integrating the expression $f(x)dx \cdot F(y)dy$ over this area, as follows.

$$P(u)\Delta u = \int_0^\infty \left[\int_{u/y}^{(u+\Delta u)/y} f(x)F(y)\, dx \right] dy$$

$$= \int_0^\infty \left[f(x')F(y) \int_{u/y}^{(u/y)+(\Delta u/y)} dx \right] dy,$$

where x' is a mean value of x between $x = u/y$ and $x = (u/y) + (\Delta u/y)$. Now at every point in the plane, $x = u/y$ (since $u = xy$). Hence we have:

$$P(u)\Delta u = \int_0^\infty [f(u'/y)F(y)(1/y)\Delta u]\, dy = \left[\int_0^\infty f(u/y)F(y)(1/y)\, dy \right] \Delta u,$$

from which the theorem follows immediately.

THEOREM 2. *Suppose a variable x is distributed in accordance with a probability law* $\int_{-\infty}^\infty f(x)dx = 1$; *and a variable y in accordance with a probability law* $\int_0^\infty F(y)dy = 1$, *x and y being independently distributed. Then the quotient,* $z = x/y$, *will be distributed according to the law* $\int_{-\infty}^\infty Q(z)dz = 1$, *where*

$$Q(z) = \int_0^\infty f(zy)F(y)y\, dy.$$

PROOF. As in the proof of Theorem 1, the (relative) number of dots in the area $dxdy$ will be $[f(x)dx][F(y)dy]$.

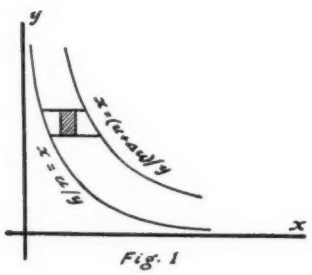

Fig. 1

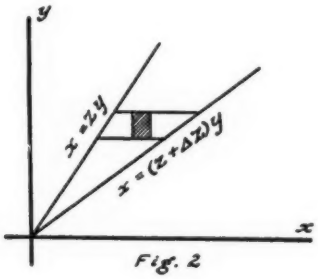

Fig. 2

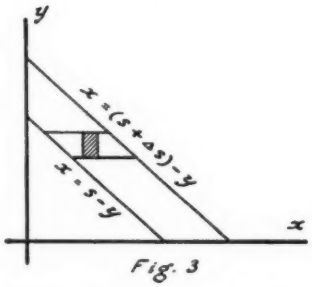

Fig. 3

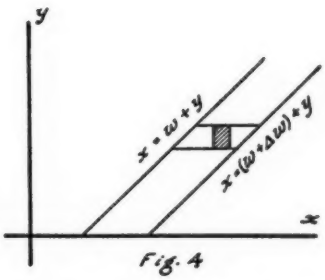

Fig. 4

Now for fixed values of z and Δz, plot the lines $x/y = z$ and $x/y = z + \Delta z$ in the xy plane, as shown in Figure 2. Then the (relative) number of dots in the area between these lines is precisely what is meant by $Q(z)\Delta z$. Hence the expression $Q(z)\Delta z$ may be built up by integrating the expression $f(x)dx \cdot F(y)dy$ over this area, as follows:

$$Q(z)\Delta z = \int_0^\infty \left[\int_{zy}^{(z+\Delta z)y} f(x)F(y)\,dx \right] dy$$

$$= \int_0^\infty \left[f(x')F(y) \int_{zy}^{zy+y\Delta z} dx \right] dy,$$

where x' is a mean value of x between $x = zy$ and $x = zy + y\Delta z$. Now at every point in the plane, $x = zy$ (since $z = x/y$). Hence we have

$$Q(z)\Delta z = \int_0^\infty [f(z'y)F(y)y\Delta z]\,dy = \left[\int_0^\infty f(zy)F(y)y\,dy \right] \Delta z,$$

from which the theorem follows immediately.

For convenience of reference, we include the corresponding theorems for the sum and difference, the proofs of which have long been well known.

THEOREM 3. *If x obeys a law $\int_0^\infty f(x)dx = 1$, and y obeys a law $\int_0^\infty F(y)dy = 1$, then the sum, $s = x + y$, will obey the law $\int_0^\infty \psi(s)ds = 1$, where*

$$\psi(s) = \int_0^\infty f(s - y)F(y)\,dy.$$

The proof consists in integrating $f(x)F(y)dxdy$ over the area bounded by the two lines $x + y = s$ and $x + y = s + \Delta s$, as shown in Figure 3.

THEOREM 4. *If x obeys a law $\int_{-\infty}^{\infty} f(x)dx = 1$, and y obeys a law $\int_{-\infty}^{\infty} F(y)dy = 1$, then the difference, $w = x - y$, will obey the law $\int_{-\infty}^{\infty} R(w)dw = 1$, where $R(w) = \int_{-\infty}^{\infty} f(w + y)\, F(y)\, dy$.*

The proof consists in integrating $f(x)F(y)dxdy$ over the area bounded by the two lines $x - y = w$ and $x - y = w + \Delta w$, as shown in Figure 4.

HARVARD UNIVERSITY.

MOMENTS ABOUT THE ARITHMETIC MEAN OF A HYPERGEOMETRIC FREQUENCY DISTRIBUTION

BY HAROLD D. LARSEN

In a recent paper[1] Kirkman has developed a method of continuation for obtaining the moments of a binomial distribution. Although other investigators[2] have found various methods which are perhaps superior from the standpoint of elegance and compactness, Kirkman's method is of some importance inasmuch as it is adaptable to use in a course in elementary statistics. With this thought in mind, we shall extend Kirkman's method to obtain the moments of the hypergeometric distribution of Table I.[3]

TABLE I

Variate v	Relative Frequency P_v
0	${}_nC_0\alpha^{(0)}\beta^{(n)}/N^{(n)}$
1	${}_nC_1\alpha^{(1)}\beta^{(n-1)}/N^{(n)}$
2	${}_nC_2\alpha^{(2)}\beta^{(n-2)}/N^{(n)}$
.	. . . . .
.	
n	${}_nC_n\alpha^{(n)}\beta^{(0)}/N^{(n)}$

[1] W. J. Kirkman, "Moments About the Arithmetic Mean of a Binomial Frequency Distribution," *Ann. Math. Statist.*, vol. vi, no. 2, June, 1935, pp. 96–101.

[2] For example, J. Riordan, "Moment Recurrence Relations for Binomial, Poisson and Hypergeometric Frequency Distributions," *Ann. Math. Statist.*, vol. viii, no. 2, June, 1937, pp. 103–111.

[3] For the Poisson distribution, this method degenerates into the application of a well-known recursion formula.

The hypergeometric distribution above can be conceived as being generated in the following manner. From an urn containing N balls, $\alpha = Np$ white and $\beta = Nq$ black, n balls are drawn without replacements. The probability that exactly v of the balls are white is

$$P_v = {}_nC_v\alpha^{(v)}\beta^{(n-v)}/N^{(n)},$$

where

$$\alpha^{(v)} = \alpha(\alpha - 1)(\alpha - 2)\cdots(\alpha - v + 1),$$

$$\alpha^{(0)} = 1, \text{ etc.}$$

It may be noted in passing that the hypergeometric distribution reduces to a binomial distribution when $n = 1$, or $N = \infty$.

For the distribution of Table I, let m_k denote the kth moment about the origin, and let μ_k denote the kth moment about the arithmetic mean. Then by definition

$$m_k = \sum_{v=0}^{n} v^k P_v,$$

and

$$\mu_k = \sum_{v=0}^{n} (v - m_1)^k P_v.$$

It is apparent that these moments are functions of the parameters α, β, n and N. In particular,

$$m_k = F(\alpha, \beta, n, N).$$

We shall have need of the hypergeometric distribution of Table II. For the latter distribution, let ν_k denote the kth moment about the origin; i. e.,

$$\nu_k = \sum_{v=0}^{n-1} v^k P'_v.$$

TABLE II

v	P'_v
0	${}_{n-1}C_0(\alpha - 1)^{(0)}\beta^{(n-1)}/(N - 1)^{(n-1)}$
1	${}_{n-1}C_1(\alpha - 1)^{(1)}\beta^{(n-2)}/(N - 1)^{(n-1)}$
2	${}_{n-1}C_2(\alpha - 1)^{(2)}\beta^{(n-3)}/(N - 1)^{(n-1)}$
.	. . . . .
.	
$n - 1$	${}_{n-1}C_{n-1}(\alpha - 1)^{(n-1)}\beta^{(0)}/(N - 1)^{(n-1)}$

Comparing Table I with Table II, we see at once that

(1) $$\nu_k = F(\alpha - 1, \beta, n - 1, N - 1).$$

In other words, ν_k is equal to the expression obtained from m_k upon replacing α, n, and N respectively by $\alpha - 1$, $n - 1$, and $N - 1$.

Now consider

$$m_k = \sum_{v=0}^{n} v^k P_v$$

$$= \sum_{v=1}^{n} v^k P_v.$$

Replacing v by $v + 1$, we have

$$m_k = \sum_{v=0}^{n-1} (v+1)^k \cdot P_{v+1}$$

$$= \frac{n\alpha}{N} \sum_{v=0}^{n-1} (v+1)^{k-1} \frac{(n-1)!}{v!(n-v-1)!} \frac{(\alpha-1)^{(v)} \beta^{(n-v-1)}}{(N-1)^{(n-1)}}$$

$$= \frac{n\alpha}{N} \sum_{v=0}^{n-1} (v+1)^{k-1} P'_v,$$

whence, expanding the binomial and summing term by term,

$$m_k = \frac{n\alpha}{N} \{\nu_{k-1} + {}_{k-1}C_1 \nu_{k-2} + {}_{k-1}C_2 \nu_{k-3} + \cdots + 1\}. \tag{2}$$

By repeated use of (1) and (2), we can obtain quite readily the moments about the origin for the distribution of Table I. It follows by definition that

$$m_0 = \sum_{v=0}^{n} P_v = 1,$$

and, similarly,

$$\nu_0 = \sum_{v=0}^{n-1} P'_v = 1.$$

Setting $k = 1$ in (2), we have

$$m_1 = \frac{n\alpha}{N} \cdot \nu_0 = n\alpha/N.$$

Setting $k = 2$, and then using (1), we obtain

$$m_2 = \frac{n\alpha}{N} \{\nu_1 + \nu_0\}$$

$$= \frac{n\alpha}{N} \left\{ \frac{(n-1)(\alpha-1)}{N-1} + 1 \right\}$$

$$= \frac{n^{(2)} \alpha^{(2)}}{N^{(2)}} + \frac{n\alpha}{N}.$$

In a similar manner,

$$m_3 = \frac{n\alpha}{N}\{\nu_2 + 2\nu_1 + \nu_0\}$$

$$= \frac{n\alpha}{N}\left\{\frac{(n-1)^{(2)}(\alpha-1)^{(2)}}{(N-1)^{(2)}} + 3\frac{(n-1)(\alpha-1)}{N-1} + 1\right\}$$

$$= \frac{n^{(3)}\alpha^{(3)}}{N^{(3)}} + 3\frac{n^{(2)}\alpha^{(2)}}{N^{(2)}} + \frac{n\alpha}{N}.$$

The coefficients are seen to follow the same law as for the binomial distribution. As a matter of fact, if we replace $\alpha^{(r)}/N^{(r)}$ by p^r in the above m's, we obtain precisely the corresponding formulae for the binomial distribution. The coefficients for some of the higher moments are

$$m_4 = \{1, 6, 7, 1\}$$

$$m_5 = \{1, 10, 25, 15, 1\}$$

$$m_6 = \{1, 15, 65, 90, 31, 1\}.$$

The moments about the arithmetic mean can now be determined from the foregoing m's by means of the semi-recursion formula

$$\mu_k = m_k - {}_kC_1\mu_{k-1}m_1 - {}_kC_2\mu_{k-2}m_1^2 - \cdots. \tag{3}$$

I have tried several formulae for this purpose, but it seems impossible to avoid a great deal of tedious reduction. Since the reduction in any case only involves algebraic manipulation, the details will be omitted. The formulae for the first few moments follow:

$$\mu_0 = 1$$

$$\mu_1 = 0$$

$$\mu_2 = npq\frac{N-n}{N-1}$$

$$\mu_3 = npq(q-p)\frac{(N-n)(N-2n)}{(N-1)(N-2)}.$$

If the higher moments are required in a practical problem, it appears to be the best course to first calculate the values of the m's, and then use (3).

THE UNIVERSITY OF NEW MEXICO.

ERRATA

The following changes should be made in my paper entitled: "On the Probability Theory of Arbitrarily Linked Events" (These *Annals*, Vol. IX, 1938):

Page 262, after 7th line from top: insert $\cdots$.

Page 262, 18th line from top: for [11] read [10].

Page 263, 5th and 6th lines from bottom: for "reasonably be assumed" read "easily be proved".

Page 264, 4th, 13th and 14th lines from top: replace each α_n by a_n .

Page 265, in (24) replace $f_{\alpha\beta}$ by $S_{\alpha\beta}$.

Page 267, 7th line from top: after $2^{2k}/(k!2^k)$ insert "(See [10], p. 13)".

Page 267, 11th line from top: for $\zeta(x)$ read $\varphi(x)$.

Page 268, 4th line from bottom: for (7) read (11).

Page 271, in (47): replace 2nd and 3rd lines by

$$-m\left[\frac{1}{\alpha}+\frac{1}{\alpha-1/n}+\cdots+\frac{1}{\alpha-(m-1)/n}\right]$$

HILDA GEIRINGER